

Tilburg University

## Exact Robust Counterparts of Ambiguous Stochastic Constraints Under Mean and Dispersion Information

Postek, Krzysztof; Ben-Tal, A.; den Hertog, Dick; Melenberg, Bertrand

*Publication date:*  
2015

*Document Version*  
Early version, also known as pre-print

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Postek, K., Ben-Tal, A., den Hertog, D., & Melenberg, B. (2015). *Exact Robust Counterparts of Ambiguous Stochastic Constraints Under Mean and Dispersion Information*. (CentER Discussion Paper; Vol. 2015-030). CentER, Center for Economic Research.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2015-030

**EXACT ROBUST COUNTERPARTS OF AMBIGUOUS  
STOCHASTIC CONSTRAINTS UNDER MEAN  
AND DISPERSION INFORMATION**

By

Krzysztof Postek, Aharon Ben-Tal,  
Dick den Hertog, Bertrand Melenberg

5 June 2015

ISSN 0924-7815  
ISSN 2213-9532

# Exact robust counterparts of ambiguous stochastic constraints under mean and dispersion information

Krzysztof Postek

CentER and Department of Econometrics and Operations Research, Tilburg University, k.postek@tilburguniversity.edu

Aharon Ben-Tal

Department of Industrial Engineering and Management, Technion - Israel Institute of Technology;  
CentER Extramural Fellow, Tilburg University, abental@ie.technion.ac.il

Dick den Hertog, Bertrand Melenberg

CentER and Department of Econometrics and Operations Research, Tilburg University,  
d.denhertog@tilburguniversity.edu, b.melenberg@tilburguniversity.edu

In this paper we consider ambiguous stochastic constraints under partial information consisting of means and dispersion measures of the underlying random parameters. Whereas the past literature used the variance as the dispersion measure, here we use the mean absolute deviation from the mean (MAD). This makes it possible to use the old result of Ben-Tal and Hochman (1972) in which *tight* upper and lower bounds on the expectation of a convex function of a random variable are given. We use these bounds to derive exact robust counterparts of expected feasibility of convex constraints and to construct new safe tractable approximations of chance constraints. Numerical examples show our method to be applicable to numerous applications of Robust Optimization, e.g., where implementation error or linear decision rules are present. Also, we show that the methodology can be used for optimization the average-case performance of worst-case optimal Robust Optimization solutions.

*Key words:* robust optimization; ambiguity; stochastic programming; chance constraints

*JEL codes:* C61

---

## 1. Introduction

Consider an optimization problem with a constraint

$$f(\mathbf{x}, \mathbf{z}) \leq 0,$$

where  $\mathbf{x} \in \mathbb{R}^{n_x}$  is the decision vector,  $\mathbf{z} \in \mathbb{R}^{n_z}$  is an uncertain parameter vector, and  $f(\cdot, \mathbf{z})$  is assumed to be convex for all  $\mathbf{z}$ . There are three principal ways to address such constraints. One of them is *Robust Optimization*. In this approach,  $\mathcal{U}$  is a user-provided convex compact uncertainty set and the constraint is to hold for all  $\mathbf{z} \in \mathcal{U}$ , i.e.,  $\mathbf{x}$  is robust feasible if:

$$\sup_{\mathbf{z} \in \mathcal{U}} f(\mathbf{x}, \mathbf{z}) \leq 0. \tag{1}$$

The key issue in this approach is to reformulate (1) to an equivalent, computationally tractable form (Ben Tal and Nemirovski (1998), Ben-Tal et al. (2009, 2015)).

In the other approaches, which go under the name of *Distributionally Robust Optimization* (DRO),  $\mathbf{z}$  is a *random* parameter vector whose distribution  $\mathbb{P}_{\mathbf{z}}$  belongs to a set  $\mathcal{P}$  (the so-called *ambiguity set*). A typical example for  $\mathcal{P}$  is a set of all distributions with given values of the first two moments. In such a setting, there are two principal constraint types: the *worst-case expected feasibility constraints*:

$$\sup_{\mathbb{P}_{\mathbf{z}} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_{\mathbf{z}}} f(\mathbf{x}, \mathbf{z}) \leq 0, \quad (2)$$

and *chance constraints*:

$$\sup_{\mathbb{P}_{\mathbf{z}} \in \mathcal{P}} \mathbb{P}_{\mathbf{z}}(f(\mathbf{x}, \mathbf{z}) > 0) \leq \epsilon. \quad (3)$$

For constraint (2) the key challenge is, for a given ambiguity set  $\mathcal{P}$ , to obtain a computationally tractable exact form of the worst-case expectation, or a good upper bound. Constraint (2) is also used in the construction of safe approximation of the ambiguous chance constraint (3), where by a *safe approximation* is meant a system  $\mathcal{S}$  of computationally tractable constraints, such that  $\mathbf{x}$  feasible for  $\mathcal{S}$  is also feasible for constraint (3).

In this paper, we consider problems with ambiguity sets consisting of distributions having given mean-dispersion measures. The literature of this type of problem started with the paper by Scarf (1958). Under mean-variance information, he derived the exact worst-case expectation formula for a single-variable piecewise linear objective function used in the newsvendor problem. Later, his result has been extended to more elaborate cases of inventory and newsvendor problems by, e.g., Gallego (1992, 2001), and Perakis and Roels (2008). Theoretical results on properties of worst-case expectation optimal solutions under general moment information were given by Rogosinsky (1958), Dupačová (1977) and Shapiro and Kleywegt (2002). Birge and Wets (1987) and Birge and Dula (1991) provide bounds on the worst-case expectations which, however, require solving an additional optimization problem and are tight only for some functions  $f(\mathbf{x}, \mathbf{z})$ .

Despite numerous works, to the best of our knowledge, no *closed-form tight upper bounds* on expectations are known for *general convex functions* in case of mean-variance information. In a recent paper by Popescu (2007), it has been proved that for a wide class of increasing concave utility functions the problem of maximizing the worst-case expected utility under mean-variance distributional information reduces to solving a parametric quadratic optimization problem.

Surprisingly, already in 1972 a result of Ben-Tal and Hochman (1972) (from now on referred to as (BH)) was available, providing exact values of tight upper and lower bounds on the expectation of convex  $f(\mathbf{x}, \cdot)$  for the case where  $\mathcal{P}$  consists of all distributions of componentwise independent  $\mathbf{z}$  with known supports, means, and mean absolute deviations from the means (MAD) of the  $z_i$ 's. In a later paper, Ben-Tal and Hochman (1985) have shown that in such a setting, tight bounds can be obtained for the optimal *decision variable* for one-dimensional problems. In this paper, we

exploit the result of (BH) to provide an exact reformulation of constraint (2) and new safe tractable approximations of constraint (3).

We choose the ambiguity setting of (BH) for the following reasons. First, the authors provide exact upper and lower bounds on the expectations of general convex  $f(\mathbf{x}, \cdot)$ . Second, the information required for their bounds, i.e., the supports, means, and MADs of the  $z_i$ 's, can easily be obtained from past data, making the method suitable for data-driven settings. The MAD measure has several desirable properties from an application's point of view, for example, its suitability to situations when the deviations of  $z_i$  are small. This and other advantages of the MAD are discussed, e.g., by El Amir (2012) and references therein.

Our contributions can be summarized as follows:

1. We propose a new method of optimizing the exact worst-case-expected performance in problems involving constraints (2) under mean-MAD partial distributional information, based on the results of (BH). We also provide a method to evaluate the exact best-case expected performance.
2. The new method is shown to be applicable to problems with constraints (2) with  $f(\mathbf{x}, \cdot)$  being either concave or convex. Concavity of  $f(\mathbf{x}, \cdot)$  is required in the classical RO framework in order to derive the tractable robust counterparts of worst-case constraints. On the other hand, convexity may appear, for example, as a result of applying linear decision rules or existence of implementation error. In the general case, convexity in the uncertain parameter leads to computationally intractable problems in the RO framework, with the exception of linear and quadratic functions. For that reason, the exact result of (BH) is particularly useful for the convex case.
3. We show that the proposed approach can be used as a second-stage method of improving the average-case performance of RO solutions, in case of existence of multiple worst-case-optimal solutions.
4. We derive new safe tractable approximations of chance constraint (3) under mean-MAD information. This is based on deriving an upper bound on the moment-generating function  $\mathbb{E}(\exp(\mathbf{w}^T \mathbf{z}))$ .
5. We show using a numerical example that optimization of the worst-case expectation of the objective function may also result in an improvement of the best-case expectation. That means, by minimizing  $\sup_{\mathbb{P}_{\mathbf{z}} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_{\mathbf{z}}} f(\mathbf{x}, \mathbf{z})$  over  $\mathbf{x}$  we are able to shift the interval for  $\mathbb{E}_{\mathbb{P}_{\mathbf{z}}} f(\mathbf{x}, \mathbf{z})$  downwards.

We mention that there are alternative ways of specifying the set  $\mathcal{P}$ , for example, as sets of distributions within some distance from a known distribution (as, for example, in Ben-Tal et al. (2013)). For a broad overview of types of ambiguity sets we refer the reader to Postek et al. (2014) and Hanasusanto et al. (2015). Some approximation results when the components of the random vector  $\mathbf{z}$  are not independent, are obtained for limited classes of function  $f(\mathbf{x}, \mathbf{z})$  in Delage and Ye (2010), Goh and Sim (2010), Calafiore and El Ghaoui (2006), and Zymmler et al. (2013). Chen et al.

(2007) propose to use so-called forward and backward deviations as characteristics of the moment generating functions of random variables to approximate chance constraints.

Wiesemann et al. (2014) have recently introduced a class of quite general ambiguity sets for which they derive computationally tractable counterparts of (2) for specific cases of  $f(\mathbf{x}, \cdot)$ . However, in their framework the components of  $\mathbf{z}$  are unrestricted in their dependence, and taking their independence into account is not straightforward. In Appendix B, we illustrate the marked difference between theirs and our robust counterparts when  $f(\mathbf{x}, \mathbf{z}) = \exp(\mathbf{x}^T \mathbf{z})$ , where without the assumption of independence one has to reformulate a robust constraint that is *strictly convex* in the uncertain parameter, requiring an exponential number of constraints.

The remainder of the paper is structured as follows. In Section 2, we describe the mean-MAD results of (BH), providing statistical background on estimation of the relevant parameters. In Section 3, we show how the mean-MAD results can be used to optimization problems involving stochastic constraints (2), including numerical examples. Section 4 includes new results on safe tractable approximations of chance constraints (3), illustrated also with a numerical study. Section 5 concludes the paper.

## 2. Bounds on the expectation of a convex function of a random variable

### 2.1. Introduction

In this section we introduce the results of (BH) on exact upper and lower bounds on the expected value of a convex function of a componentwise independent  $\mathbf{z} = (z_1, \dots, z_{n_{\mathbf{z}}})^T$ . From now on we drop the subscript  $\mathbf{z}$  from  $\mathbb{P}_{\mathbf{z}}$  and the probability distribution applies to  $\mathbf{z}$ . The pieces of partial distributional information on  $z_i$ 's constituting the ambiguity sets in (BH) are:

- (i) support including intervals:  $\text{supp}(z_i) \subseteq [a_i, b_i]$ , where  $-\infty < a_i \leq b_i < \infty, i = 1, \dots, n_{\mathbf{z}}$ . (BH) show also that their bounds hold in cases where  $a_i = -\infty$  and/or  $b_i = +\infty$ . We illustrate this in Remark 3. In the remainder of the paper, however, we concentrate on the bounded case, with RO applications in mind.
- (ii) means:  $\mathbb{E}_{\mathbb{P}}(z_i) = \mu_i$ ,
- (iii) mean absolute deviations from the means (MAD):  $\mathbb{E}_{\mathbb{P}}|z_i - \mu_i| = d_i$ . The MAD is known to satisfy the bound ((BH), Lemma 1):

$$0 \leq d_i \leq d_{i,\max} = \frac{2(b_i - \mu_i)(\mu_i - a_i)}{(b_i - a_i)}, \quad i = 1, \dots, n_{\mathbf{z}}, \quad (4)$$

- (iv) probabilities of  $z_i$ 's being greater than or equal to  $\mu_i$ :  $\mathbb{P}(z_i \geq \mu_i) = \beta_i$ . For example, in the case of continuous symmetric distribution of  $z_i$  we know that  $\beta_i = 0.5$ . This quantity is known to satisfy the bounds:

$$\frac{d_i}{2(b_i - \mu_i)} = \underline{\beta}_i \leq \beta_i \leq \bar{\beta}_i = 1 - \frac{d_i}{2(\mu_i - a_i)}, \quad i = 1, \dots, n_{\mathbf{z}}. \quad (5)$$

Using these building blocks, we define two types of ambiguity set  $\mathcal{P}$ :

- the  $(\mu, d)$  ambiguity set, consisting of the distributions with known (i), (ii), and (iii) for each  $z_i$ :

$$\mathcal{P}_{(\mu, d)} = \{\mathbb{P} : \text{supp}(z_i) \subseteq [a_i, b_i], \quad \mathbb{E}_{\mathbb{P}}(z_i) = \mu, \quad \mathbb{E}_{\mathbb{P}}|z_i - \mu_i| = d_i, \quad \forall i, \quad z_i \perp z_j, \quad \forall i \neq j\}, \quad (6)$$

where  $z_i \perp z_j$  denotes the stochastic independence of components  $z_i$  and  $z_j$ ,

- the  $(\mu, d, \beta)$  ambiguity set, consisting of the distributions with known (i), (ii), (iii), and (iv) for each  $z_i$ :

$$\mathcal{P}_{(\mu, d, \beta)} = \{\mathbb{P} : \mathbb{P} \in \mathcal{P}_{(\mu, d)}, \quad \mathbb{P}(z_i \geq \mu_i) = \beta_i, \quad \forall i\}. \quad (7)$$

In the following, we present the results of (BH) on  $\max_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z})$ ,  $\max_{\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z})$  and  $\min_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z})$ ,  $\min_{\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z})$ , where  $f : \mathbb{R}^{n_z} \mapsto \mathbb{R}$  is convex. We note that in the case of concave  $f(\cdot)$  the upper bounds become lower bounds and vice versa.

## 2.2. One-dimensional $z$

We begin with the simpler and more illustrative case of one-dimensional random variable  $z$ . For that reason, we drop the subscript  $i$ .

**Upper bounds.** (BH) shows that:

$$\max_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(z) = p_1 f(a) + p_2 f(\mu) + p_3 f(b), \quad (8)$$

where:

$$p_1 = \frac{d}{2(\mu - a)}, \quad p_2 = 1 - \frac{d}{2(\mu - a)} - \frac{d}{2(b - \mu)}, \quad p_3 = \frac{d}{2(b - \mu)}. \quad (9)$$

Hence, the worst-case distribution is a three-point distribution on  $\{a, \mu, b\}$ . The same bound holds for the  $(\mu, d, \beta)$  ambiguity set.

**REMARK 1.** A special case of (9) is the upper bound on  $f(z)$  when only the interval  $[a, b]$  and the mean  $\mu$  are known. Such a bound is known as the Edmundson-Madansky bound (Edmundson 1956, Madansky 1959):

$$\max_{\mathbb{P} \in \mathcal{P}_{(\mu)}} f(z) = \frac{b - \mu}{b - a} f(a) + \frac{\mu - a}{b - a} f(b) \text{ where } \mathcal{P}_{(\mu)} = \{\mathbb{P} : \text{supp}(z) \subseteq [a, b], \quad \mathbb{E}_{\mathbb{P}} z = \mu\}. \quad (10)$$

Indeed, inserting the biggest possible value of MAD (see (4)) equal to  $d_{\max} = 2(b - \mu)(\mu - a)/(b - a)$  into (9) yields the probability of outcome  $\mu$  equal to 0.  $\square$

**Lower bounds.** To obtain a closed-form lower bound on  $\mathbb{E}_{\mathbb{P}} f(z)$ , additional information is needed in the form of the parameter  $\beta$ . Then, it holds that:

$$\min_{\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}} \mathbb{E}_{\mathbb{P}} f(z) = \beta f\left(\mu + \frac{d}{2\beta}\right) + (1 - \beta) f\left(\mu - \frac{d}{2(1 - \beta)}\right). \quad (11)$$

In case  $\beta$  is not known, (BH) shows that:

$$\min_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(z) = \min_{\underline{\beta} \leq \beta \leq \bar{\beta}} \left\{ \beta f\left(\mu + \frac{d}{2\beta}\right) + (1 - \beta) f\left(\mu - \frac{d}{2(1 - \beta)}\right) \right\}, \quad (12)$$

where the minimization over  $\beta$  is a convex problem in  $\beta$  and for a strictly convex  $f(\cdot)$  there is a unique optimal solution.

REMARK 2. In case of no knowledge about  $d$ , the lower bound is obtained at  $d^* = 0$ , which corresponds to the well-known Jensen bound (Jensen 1906).  $\square$

REMARK 3. In case where  $a = -\infty$  and/or  $b = +\infty$ , bounds can still be obtained under additional conditions, namely that the limits  $\lim_{t \rightarrow \pm\infty} f(t)/t$  exist and are finite, with the ‘+’ corresponding to  $b = +\infty$ , and the ‘−’ corresponding to  $a = -\infty$ . We illustrate this on the example  $a \in \mathbb{R}, b = +\infty$ . Assume that  $\lim_{t \rightarrow +\infty} f(t)/t = \gamma$ . We then have:

$$\begin{aligned} \max_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(z) &= \max_{\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}} \mathbb{E}_{\mathbb{P}} f(z) = \lim_{b \rightarrow \infty} \left\{ \frac{d}{2(\mu - a)} f(a) + \left(1 - \frac{d}{2(\mu - a)} - \frac{d}{2(b - \mu)}\right) f(\mu) + \frac{d}{2(b - \mu)} f(b) \right\} \\ &= \frac{d}{2(\mu - a)} f(a) + \left(1 - \frac{d}{2(\mu - a)}\right) f(\mu) + \frac{d}{2} \gamma, \end{aligned}$$

and for the lower bound we have:

$$\min_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} f(z) = \frac{d}{2} \gamma + f\left(\mu - \frac{d}{2}\right).$$

The lower bound for the  $(\mu, d, \beta)$  ambiguity set is the same as (11).  $\square$

### 2.3. Multidimensional $\mathbf{z}$

**Upper bounds.** For  $n_{\mathbf{z}} > 1$ , the worst-case probability distribution under  $(\mu, d)$  information is a componentwise counterpart of (9):

$$p_1^i = \frac{d_i}{2(\mu_i - a_i)}, \quad p_2^i = 1 - \frac{d_i}{2(\mu_i - a_i)} - \frac{d_i}{2(b_i - \mu_i)}, \quad p_3^i = \frac{d_i}{2(b_i - \mu_i)}, \quad i = 1, \dots, n_{\mathbf{z}}. \quad (13)$$

The worst-case expectation of  $f(\mathbf{z})$  is obtained by applying the bound (8) for each  $z_i$ , i.e., by enumerating over all  $3^{n_{\mathbf{z}}}$  permutations of outcomes  $a_i, \mu_i, b_i$  of components  $z_i$ . It holds then that (BH):

$$\max_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}) = \sum_{\alpha \in \{1, 2, 3\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} p_{\alpha_i}^i f(\tau_{\alpha_1}^1, \dots, \tau_{\alpha_{n_{\mathbf{z}}}}^{n_{\mathbf{z}}}), \quad (14)$$

where

$$\tau_1^i = a_i, \quad \tau_2^i = \mu_i, \quad \tau_3^i = b_i \quad \text{for } i = 1, \dots, n_{\mathbf{z}}. \quad (15)$$



Again, the same upper bound holds for the  $(\mu, d, \beta)$  ambiguity set.

**Lower bounds.** Similar to the one-dimensional case, the closed-form lower bound under  $(\mu, d)$  information requires known  $\beta = (\beta_1, \dots, \beta_{n_z})^T$ :

$$\min_{\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}) = \sum_{\alpha \in \{1, 2\}^{n_z}} \prod_{i=1}^{n_z} q_{\alpha_i}^i f(v_{\alpha_1}^1, \dots, v_{\alpha_{n_z}}^{n_z}), \quad (16)$$

where  $\underline{\beta} = (\underline{\beta}_1, \dots, \underline{\beta}_{n_z})^T$ ,  $\bar{\beta} = (\bar{\beta}_1, \dots, \bar{\beta}_{n_z})^T$  and

$$q_1^i = \beta_i, \quad q_2^i = 1 - \beta_i, \quad v_1^i = \mu_i + d_i/2\beta_i, \quad v_2^i = \mu_i - d_i/2(1 - \beta_i). \quad (17)$$

If  $\beta$  is unknown, the bound is obtained by minimization:

$$\min_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z}) = \inf_{\underline{\beta} \leq \beta \leq \bar{\beta}} \sum_{\alpha \in \{1, 2\}^{n_z}} \prod_{i=1}^{n_z} q_{\alpha_i}^i f(v_{\alpha_1}^1, \dots, v_{\alpha_{n_z}}^{n_z}). \quad (18)$$

In the multidimensional case, minimization over  $\beta$  is a nonconvex problem - it is only convex in  $\beta_i$  when other  $\beta_j, j \neq i$  are fixed.

## 2.4. Estimating $\mu$ , $d$ , and $\beta$

As the bounds on the expectation of a random variable depend on the parameters  $a$ ,  $b$ ,  $\mu$ ,  $d$ , and  $\beta$ , it is necessary to know or estimate these parameters, and decide ‘how much information is actually available’. Here, we provide the reader with a simple procedure to achieve this. We operate here with the one-dimensional case for  $z$ , and the multi-dimensional case follows straightforwardly due to the independence of the components of  $\mathbf{z}$ . Appendix C describes the properties of the MAD in relation to the variance and formulas for the MAD of several important classes of probability distribution.

First, we introduce estimators of  $\mu$ ,  $d$ , and  $\beta$  and discuss their asymptotic properties. Based on these results, we provide a procedure that can be used to assess whether the amount of information available is sufficient to use the results for the  $(\mu, d)$  ambiguity set or the  $(\mu, d, \beta)$  ambiguity set.

Let  $z^{(1)}, \dots, z^{(n)}$  be a random sample of the values of  $z$ . We assume the interval  $[a, b]$  to be fixed by the user. As estimators for  $\mu$ ,  $d$ , and  $\beta$  we consider

- $\hat{\mu} = \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$ , the sample mean;
- $\hat{d} = \frac{1}{n} \sum_{i=1}^n |z_i - \bar{z}|$ , the sample MAD;
- $\hat{\beta} = \frac{1}{n} \sum_{i=1}^n 1_{\bar{z}, \infty}(z_i)$ , the sample analogue of  $\beta$ .

Let  $\hat{\theta} = (\hat{\mu}, \hat{d}, \hat{\beta})^T$  and  $\theta = (\mu, d, \beta)^T$ . Then we have

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n} \frac{1}{n} \sum_{i=1}^n \tilde{\psi}(z_i) + o_p(1),$$

with  $\tilde{\psi}(z) = (\tilde{\psi}_\mu(z), \tilde{\psi}_d(z), \tilde{\psi}_\beta(z))^T$  defined by

$$\begin{aligned}\tilde{\psi}_\mu(z) &= z - \mu, \\ \tilde{\psi}_d(z) &= 2 \left( (z - \mu) + ([z1_{(\mu, \infty)}(z) - z\beta] - \frac{1}{2}d) - \mu(1_{(\mu, \infty)}(z) - \beta) \right), \\ \tilde{\psi}_\beta(z) &= (1_{(\mu, \infty)}(z) - \beta) - (z - \mu)p(\mu),\end{aligned}$$

where  $p(\mu)$  stands for the density function of  $z$  evaluated at  $\mu$ , assuming that  $\mathbb{P}$  represents a continuous distribution (in which case  $p(\cdot)$  is assumed to be continuous in a neighborhood of  $\mu$ ). The expression for  $\tilde{\psi}_\mu(z)$  is standard. The expression  $\tilde{\psi}_d(z)$  is based on Gastwirth (1974). The expression  $\tilde{\psi}_\beta(z)$  follows from arguments presented in Gastwirth (1974). As a consequence, we find for the limit distribution of  $\hat{\theta}$ :

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow_d N(0, \text{cov}(\tilde{\psi})). \quad (19)$$

The asymptotic covariance matrix  $\text{cov}(\tilde{\psi}) = E(\tilde{\psi}(z)\tilde{\psi}(z)^T)$  can be estimated consistently by

$$\widehat{\text{cov}}(\psi) = \frac{1}{n} \sum_{i=1}^n \hat{\tilde{\psi}}(z_i) \hat{\tilde{\psi}}(z_i)^T,$$

with  $\hat{\tilde{\psi}}(z_i)$  obtained from  $\tilde{\psi}(z_i)$  by replacing  $\mu$ ,  $d$ , and  $\beta$  by their estimates  $\hat{\mu}$ ,  $\hat{d}$ , and  $\hat{\beta}$ , and where  $p(\cdot)$  is replaced by some (appropriately chosen) consistent estimator  $\hat{p}(\cdot)$ .

We now proceed to the proper estimation of the parameters of distribution of  $z$ . The parameters satisfy the bounds

$$a \leq \mu \leq b, \quad 0 \leq d \leq d_{\max}, \quad \underline{\beta} \leq \beta \leq \bar{\beta},$$

with

$$r_\beta = \bar{\beta} - \underline{\beta} = 4 - \frac{1}{2} \frac{d(b-a)}{(\mu-a)(b-\mu)}.$$

We can estimate  $d_{\max}$  consistently by  $\hat{d}_{\max}$  (by estimating  $\mu$  by  $\hat{\mu}$ ) and  $r_\beta$  consistently by  $\hat{r}_\beta$  (by estimating  $\mu$  by  $\hat{\mu}$  and  $d$  by  $\hat{d}$ ). If  $\hat{d}_{\max}$  is not significantly different from 0, then there is not much empirical support for assuming that we ‘know’  $d$ . Similarly, if  $\hat{r}_\beta$  is not significantly different from 0, then there is not much empirical support for assuming that we ‘know’  $\beta$ . The (asymptotic) accuracy of  $\hat{d}_{\max}$  and  $\hat{r}_\beta$  can easily be quantified using the ‘delta method’, resulting in  $\sqrt{n}(\hat{d}_{\max} - d_{\max}) \rightarrow_d N(0, \sigma_{d_{\max}}^2)$  and  $\sqrt{n}(\hat{r}_\beta - r_\beta) \rightarrow_d N(0, \sigma_{r_\beta}^2)$ .<sup>1</sup> With these definitions, we present now our procedure for estimation of the information basis for the use of the bounds:

<sup>1</sup>The ‘delta method’ yields  $\sigma_{d_{\max}}^2 = r^2 \text{var}(\tilde{\psi}_\mu)$ , with  $r = \frac{\partial d_{\max}}{\partial \mu} = \frac{2(b+a-2\mu)}{b-a}$ . Similarly, we have  $\sigma_{r_\beta}^2 = s^T \text{cov}((\tilde{\psi}_\mu, \tilde{\psi}_d)^T) s$ , with  $s = \frac{\partial r_\beta}{\partial (\mu, d)^T} = \left( -\frac{d(b-a)(b+a-2\mu)}{2((\mu-a)(b-\mu))^2}, -\frac{b-a}{2(\mu-a)(b-\mu)} \right)^T$ .

1. Estimate  $\mu$  by  $\hat{\mu}$ , and quantify the accuracy of the latter (using the limit distribution given in (19)). Decide whether the accuracy is high enough to proceed under the assumption of a ‘known’  $\mu$ . If so, go to step 2.
2. Test the hypothesis  $H_0 : d_{\max} = 0$  against  $H_1 : d_{\max} > 0$ , using as test statistic  $\hat{d}_{\max} / \sqrt{\hat{\sigma}_{d_{\max}}/n}$ . This is a one-sided test. If  $H_0$  is rejected ( $H_1$  accepted), go to step 3.
3. Estimate  $d$  by  $\hat{d}$ , and quantify the accuracy of the latter (using the limit distribution given in (19)). Decide whether the accuracy is high enough to proceed under the assumption of a ‘known’  $d$ . If so, go to step 4.
4. Test the hypothesis  $H_0 : r_\beta = 0$  against  $H_1 : r_\beta > 0$ , using as test statistic  $\hat{r}_\beta / \sqrt{\hat{\sigma}_{r_\beta}/n}$ . This is a one-sided test. If  $H_0$  is rejected ( $H_1$  accepted), go to step 5.
5. Estimate  $\beta$  by  $\hat{\beta}$ , and quantify the accuracy of the latter (using the limit distribution given in (19)). Decide whether the accuracy is high enough to proceed under the assumption of a ‘known’  $\beta$ .

It may turn out that credible information is available only about the support, or support and the mean of  $z$ . In the first case, when only the support-including interval  $[a, b]$  is known, a larger sample is needed to estimate other parameters. In case the support  $[a, b]$  and  $\mu$  are known, one may use the results of Edmundson-Madansky for the upper bound (see Remark 1) and Jensen for the lower bounds (see Remark 2).

### 3. Robust counterparts of expected feasibility constraints

In this section we demonstrate how the results of (BH) can be used to reformulate problems involving worst-case expected feasibility constraints:

$$\max_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z}) \leq 0 \quad \rightarrow \quad g(\mathbf{x}) \leq 0, \quad (20)$$

where  $f(\cdot, \mathbf{z})$  is convex for all  $\mathbf{z}$ , by providing explicit forms for the inner maximization over  $\mathcal{P} = \mathcal{P}_{(\mu, d)}$  and  $\mathcal{P} = \mathcal{P}_{(\mu, d, \beta)}$  using the results of (BH). The resulting forms can also be used to evaluate the minimum expected value of  $f(\mathbf{x}, \mathbf{z})$  after the optimal solution  $\mathbf{x}$  is found. We consider two cases, depending on the convexity/concavity of  $f(\mathbf{x}, \cdot)$ . First, we show how the corresponding results can be used to reformulate a single constraint. Later, we introduce three particular situations in which convexity/concavity in the uncertain parameter may occur. In the end, we give numerical applications of the proposed methodology.

#### 3.1. Reformulating the constraints - convex case

Consider the case where  $f(\mathbf{x}, \cdot)$  in (20) is convex for all  $\mathbf{x}$ . By constructing the upper bound (14), inequality (20) reduces to (where the subindex  $U$  refers to the upper bound):

$$g_U(\mathbf{x}) = \sum_{\alpha \in \{1, 2, 3\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} p_{\alpha_i}^i f(\mathbf{x}, \tau_{\alpha_1}^1, \dots, \tau_{\alpha_{n_{\mathbf{z}}}}^{n_{\mathbf{z}}}) \leq 0, \quad (21)$$

with  $p_{\alpha_i}^i, \tau_{\alpha_i}^i$  defined as in (13) and (15). As we can see,  $g_U(\cdot)$  in (21) inherits the convexity in  $\mathbf{x}$  from  $f(\cdot, \mathbf{z})$  and its computational complexity is dependent only on the complexity of  $f(\cdot, \mathbf{z})$ .

### 3.2. Reformulating the constraints - concave case

Consider now the case where  $f(\mathbf{x}, \cdot)$  in (20) is concave for all  $\mathbf{x}$ . Since (18) is a lower bound for convex  $f(\mathbf{x}, \cdot)$ , we can use it to obtain an upper bound on the concave function. Thus, (20) is equivalent to (where the subindex  $L$  refers to the lower bound):

$$g_L(\mathbf{x}) = \sum_{\alpha \in \{1,2\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} q_{\alpha_i}^i f(\mathbf{x}, v_{\alpha_1}^1, \dots, v_{\alpha_{n_{\mathbf{z}}}}^{n_{\mathbf{z}}}) \leq 0, \quad (22)$$

with  $q_{\alpha_i}^i, v_{\alpha_i}^i$  defined by (17). In case  $\beta$  is unknown, constraint (20) is equivalent to:

$$\sup_{\underline{\beta} \leq \beta \leq \bar{\beta}} g_L(\mathbf{x}) \leq 0, \quad (23)$$

with  $\underline{\beta}, \bar{\beta}$  defined in (5). In general, the maximization on the left-hand side is a nonconvex problem in  $\beta$ . One way to handle such a constraint is to estimate the value of  $\beta$  and insert it into the left-hand side. Alternatively, one can notice that  $g_L(\mathbf{x})$  is convex in  $\beta_i$  when all  $\beta_j, j \neq i$  are fixed. Then, one can apply the so-called *adversarial approach*, that is, iteratively solving the problem involving the ambiguous constraint and searching for values of  $\beta$  making the constraint infeasible.

### 3.3. Evaluation of the lower bound on the expectation

Once an optimal solution  $\bar{\mathbf{x}}$  is found, the decision maker may be interested not only in the upper bound on  $\mathbb{E}_{\mathbb{P}} f(\bar{\mathbf{x}}, \mathbf{z})$  but also on its lower bound. This can be the case particularly when  $f(\bar{\mathbf{x}}, \mathbf{z})$  represents the objective function. Then, knowledge on the upper and lower bounds provides an interval to which the expected value of the objective belongs.

For that purpose, the results of (BH) on upper bounds for convex functions can be used to obtain lower bounds on concave functions and vice versa. In the concave case, exact lower bound is then given by  $g_U(\bar{\mathbf{x}})$ . In the convex case, the lower bound is given by  $g_L(\bar{\mathbf{x}})$  or  $\inf_{\underline{\beta} \leq \beta \leq \bar{\beta}} g_L(\bar{\mathbf{x}})$ , if  $\beta$  is unknown.

### 3.4. Dependent random parameters and dimensionality

In this section we discuss how to address the potential difficulties facing our DRO approach - (i) possible dependence between the components of  $\mathbf{z}$  and (ii) the dimensionality of  $\mathbf{z}$ .

**Dependence.** If the random uncertain vector  $\mathbf{z}$  contains dependent components, they can be decomposed by means of factor analysis, for example, based on Principal Component Analysis (see Jolliffe (2002)), into linear combinations of a limited number of uncorrelated factors. For example,

in a situation of portfolio optimization problem with 25 assets, it is natural to decompose them into 3-4 uncorrelated risk factors, whose empirical distribution provides information also about their support, means and MADs. Even though uncorrelatedness can be much weaker than independence, such a technique is often a practical solution.

**Dimensionality.** The functions  $g_U(\cdot)$  and  $g_L(\cdot)$  include  $3^{n_z}$  and  $2^{n_z}$  terms, respectively, which can be prohibitively large. However, there are specific cases where this difficulty may be alleviated and we provide two such examples in the following.

- **$f(\mathbf{x}, \mathbf{z})$  is a sum of functions with smaller number of arguments.** Let:

$$f(\mathbf{x}, \mathbf{z}) = \sum_{j=1}^{n_c} f^{(j)}(\mathbf{x}, \mathbf{z}^{(j)}),$$

where  $\mathbf{z}^{(j)}, j = 1, \dots, n_c$  are (possibly overlapping) subvectors of  $\mathbf{z}$ , with  $\mathbf{z}^{(j)} \in \mathbb{R}^{n_j}$ . Then,

$$\sup_{\mathbb{P} \in \mathcal{P}} f(\mathbf{x}, \mathbf{z}) = \mathbb{E}_{\mathbb{P}^*_{(\mu, d)}} \left( \sum_{j=1}^{n_c} f^{(j)}(\mathbf{x}, \mathbf{z}^{(j)}) \right) = \sum_{j=1}^{n_c} \mathbb{E}_{\mathbb{P}^*_{(\mu, d)}} f^{(j)}(\mathbf{x}, \mathbf{z}^{(j)}).$$

In this way, it is necessary to evaluate only the expectations of  $f^{(j)}(\mathbf{x}, \mathbf{z}^{(j)})$  separately and one obtains a summation of only  $\sum_{j=1}^{n_c} 3^{|n_j|}$  terms in the  $(\mu, d)$  case.

Moreover, in some cases, after all the worst-case expectations of  $f^{(j)}(\mathbf{x}, \mathbf{z}^{(j)})$  are formulated as functions of  $a_i, b_i, \mu_i$ , and  $d_i$ , the resulting formulas can be greatly simplified. That is, terms involving the same functions of  $\mathbf{x}$ , coming from different worst-case expectations, can be added to each other. This is the case, for example, when  $f(\mathbf{x}, \mathbf{z})$  is polynomial in  $\mathbf{x}$ .

- **Moment generating functions.** An important special case is the function  $f(\mathbf{x}, \mathbf{z}) = \exp(\mathbf{x}^T \mathbf{z})$ . Upper bounds on moment generating functions  $\mathbb{E} \exp(\mathbf{x}^T \mathbf{z})$  are a key tool in constructing safe tractable approximations of chance constraints. As we show in Section 4, the properties of the  $\exp(\cdot)$  allow for a simple, closed-form formula for its worst-case expectation under  $(\mu, d)$  information and for which the number of terms is small.

### 3.5. The use of the (BH) bounds in some general applications

In this section we present three cases where the reformulations of the worst-case expected feasibility constraints presented in Sections 3.1 and 3.2 can be used.

**Average-case enhancement of RO solutions.** The first application lies in finding worst-case-optimal solutions with good average-case performance to the following RO problem:

$$\begin{aligned} \min_{\mathbf{x}, t} \quad & t \\ \text{s.t.} \quad & \sup_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{x}, \mathbf{z}) \leq t, \\ & \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\mathbf{x}, \mathbf{z}) \leq 0, \quad i = 1, \dots, m. \end{aligned} \tag{24}$$

It happens frequently that there exist multiple optimal solutions to (24), see Iancu and Trichakis (2013), de Ruiter et al. (2014). Whereas the worst-case performance of such solutions is the same, their average-case performance may differ dramatically. For that reason, once the optimal value  $\bar{t}$  for (24) is known, a second optimization step may be used to select one of the optimal solutions to provide good average-case behavior. Since the results of (BH) provide exact bounds on the worst-case expectations, they can be used in such a step. In the following, we describe such a two-step procedure:

1. Solve problem (24) and denote its optimal value by  $\bar{t}$ .
2. Solve the following problem, minimizing the worst-case expectation of the objective value, with the worst-case value of  $f(\mathbf{x}, \mathbf{z})$  less than or equal to  $\bar{t}$ :

$$\begin{aligned}
 & \min_{\mathbf{x}, t} u \\
 & \text{s.t. } \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z}) \leq u \\
 & \quad \sup_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{x}, \mathbf{z}) \leq \bar{t}, \\
 & \quad \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\mathbf{x}, \mathbf{z}) \leq 0, \quad i = 1, \dots, m.
 \end{aligned} \tag{25}$$

In case of multiple optimal solutions to (24), the two-step procedure is expected to select the optimal solution with good average-case performance for its focus on the worst-case expectation among the best worst-case solutions. If the uncertainty is present only in the constraints involving functions  $g_i(\cdot, \cdot)$ , a similar two-step approach can be designed to maximize the worst-case expected slack in the worst-case constraints in (24), see Iancu and Trichakis (2013).

An alternative approach to enhancing robust solutions is to sample a number  $S$  of scenarios for  $\mathbf{z}$  to find a solution that optimizes the average of the objective value over the sample.<sup>2</sup> This approach, however, has as shortcoming that the outcome might depend on the choice of sample size  $S$  and the sample itself. For that reason, the DRO methods can provide a good alternative to enhancing the quality of RO solutions. In our paper, we test the application of the  $(\mu, d)$  bounds to an inventory management problem in Section 3.7.

**Implementation error.** The second application we consider is when the decision variables cannot be implemented with the designed value due to implementation error in the following problem:

$$\begin{aligned}
 & \min_{\mathbf{x}, t} t \\
 & \text{s.t. } f(\mathbf{x}) \leq t, \\
 & \quad g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m.
 \end{aligned} \tag{26}$$

In case of the existence of an additive implementation error  $\mathbf{z}$  the implemented value is  $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{z}$ , where  $\bar{\mathbf{x}}$  is the designed value and  $\mathbf{z} = (z_1, \dots, z_{n_{\mathbf{x}}})^T$  is the error. Then, the problem becomes:

$$\begin{aligned}
 & \min_{\bar{\mathbf{x}}, t} t \\
 & \text{s.t. } f(\bar{\mathbf{x}} + \mathbf{z}) \leq t, \\
 & \quad g_i(\bar{\mathbf{x}} + \mathbf{z}) \leq 0, \quad i = 1, \dots, m.
 \end{aligned} \tag{27}$$

<sup>2</sup> As a special case, one can choose only one scenario, corresponding to the nominal values of the uncertain parameters.

Since  $f(\mathbf{x})$  is convex in  $\mathbf{x}$ , in (27) the function  $f(\bar{\mathbf{x}} + \mathbf{z})$  is convex in  $\mathbf{z}$ . For that reason, optimization of the worst-case value of the objective function could be difficult, as typically RO techniques rely on the constraint being concave in the uncertain parameter (see Ben-Tal et al. (2009, 2015)). Therefore, optimizing the worst-case values of convex constraints under implementation error is a problem leading to computational intractability, apart from special cases such as linear constraints (see Ben-Tal et al. (2015)) or (conic) quadratic constraints with simultaneously diagonalizable quadratic forms defining the constraint and the uncertainty set for the error (see Ben-Tal and den Hertog (2011)).

Because of the above, it may be an alternative to optimize the worst-case expectation of the objective function, for which our DRO method applies under the corresponding distributional assumptions on  $\mathbf{z}$ , i.e., that the ambiguity set for the distribution of  $\mathbf{z}$  is  $\mathcal{P}_{(\mu, d)}$ . Then, the problem becomes:

$$\begin{aligned} \min_{\bar{\mathbf{x}}, t} \quad & t \\ \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\bar{\mathbf{x}} + \mathbf{z}) \leq t, \\ & \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\bar{\mathbf{x}} + \mathbf{z}) \leq 0, \quad i = 1, \dots, m. \end{aligned} \quad (28)$$

The first constraint in (28) is convex in  $\mathbf{z}$  and one can apply the reformulation (21). Similarly, one can reformulate a problem where multiplicative error occurs, i.e., where  $\mathbf{x} = (\bar{x}_1 z_1, \dots, \bar{x}_{n_x} z_{n_x})^T$ .

**Convex constraints and linear decision rules.** The third application of our DRO approach comes when the constraints of a problem are convex in  $z$  as a result of applying linear decision rules. To show how such a situation occurs, we consider a two-stage RO problem:

$$\begin{aligned} \min_{\mathbf{x}_1, \mathbf{x}_2, t} \quad & t \\ \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}_1, \mathbf{x}_2(\mathbf{z}), \mathbf{z}) \leq t \\ & \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\mathbf{x}_1, \mathbf{x}_2(\mathbf{z}), \mathbf{z}) \leq 0, \quad i = 1, \dots, m, \end{aligned} \quad (29)$$

where  $\mathbf{x}_1 \in \mathbb{R}^{n_{x1}}$  is implemented before  $\mathbf{z}$  is known and  $\mathbf{x}_2 \in \mathbb{R}^{n_{x2}}$  is implemented after  $\mathbf{z}$  is known, i.e.  $\mathbf{x}_2 = \mathbf{x}_2(z)$ . In such cases, it is possible to define the time-2 decisions as a linear function  $\mathbf{x}_2(z) = \mathbf{v} + \mathbf{V}\mathbf{z}$  of the uncertain parameter  $z$  (see Ben-Tal et al. (2004)), to provide adjustability of decisions at time 2.<sup>3</sup> The problem is then:

$$\begin{aligned} \min_{\mathbf{x}_1, \mathbf{v}, \mathbf{V}, t} \quad & t \\ \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}_1, \mathbf{v} + \mathbf{V}\mathbf{z}, \mathbf{z}) \leq t \\ & \sup_{\mathbf{z} \in \mathcal{Z}} g_i(\mathbf{x}_1, \mathbf{v} + \mathbf{V}\mathbf{z}, \mathbf{z}) \leq 0, \quad i = 1, \dots, m. \end{aligned} \quad (30)$$

Since  $f(\mathbf{x}_1, \mathbf{x}_2(\mathbf{z}), \mathbf{z})$  is convex in  $\mathbf{x}_2$ , the first constraint in (30) may also be convex in  $\mathbf{z}$ . In such a case, a further reformulation of problem (29) can be conducted as in Section 3.5. We combine linear decision rules with  $(\mu, d)$  information in the inventory problem of Sections 3.6 and 3.7.

<sup>3</sup> One may also use other decision rules. However, we limit ourselves only to the analysis of the linear case as the linear decision rules are very often a powerful enough tool, see Bertsimas et al. (2011). Moreover, the (non)convexity of the problem resulting from application of linear decision rules is easy to verify, see Boyd and Vandenberghe (2004).

### 3.6. Application 1: Inventory management - average case performance

**Introduction.** In this section we consider an application of the  $(\mu, d)$  method to minimization of the average-case costs in inventory management. The main research questions are:

1. How does minimizing the worst-case expectation affect the best-case expectation under the given distributional assumptions?
2. What is the average-case performance of solutions minimizing the worst-case expectation compared to the adjustable robust solutions, minimizing the worst-case outcome of the objective values?

To answer them, we adapt the numerical example from Ben-Tal et al. (2005) with a single product and where inventory is managed periodically over  $T$  periods. At the beginning of each period  $t$  the decision maker has an inventory of size  $x_t$  and he orders a quantity  $q_t$  for unit price  $c_t$ . The customers then place their demands  $z_t$ . The retailer's status at the beginning of the planning horizon is given through the parameter  $x_1$  (initial inventory). Apart from the ordering costs, the following costs are incurred over the planning horizon:

- holding cost  $h_t \max\{0, x_t + q_t - z_t\}$ , where  $h_t$  are the unit holding costs,
- shortage costs  $p_t \max\{0, z_t - x_t - q_t\}$ , where  $p_t$  are the unit shortage costs.

Inventory  $x_{T+1}$  left at the end of period  $T$  has a unit salvage value  $s$ . Also, one must impose  $h_T - s \geq -p_T$  to maintain the problem's convexity. The constraints in the model include:

- balance equations linking the inventory in each period to the inventory, order quantity, and demand in the preceding period,
- upper and lower bounds on the order quantities in each period  $L_t \leq q_t \leq U_t$ ,
- upper and lower bounds on cumulative order quantities in each period  $\hat{L}_t \leq \sum_{\tau=1}^t q_\tau \leq \hat{U}_t$ .

With ordering decisions  $\mathbf{q}(\mathbf{z}) = (q_1, q_2(\mathbf{z}^1), \dots, q_T(\mathbf{z}^{T-1}))^T$ , where  $\mathbf{z}^t = (z_1, \dots, z_t)^T$ , the objective function value for a given demand vector  $\mathbf{z}$  is

$$f(\mathbf{q}(\mathbf{z}), \mathbf{z}) = -s \max\{x_{T+1,0}(\mathbf{z}^T)\} + \sum_{t=1}^T (c_t q_t(\mathbf{z}^{t-1}) + h_t \max\{x_{t+1}(\mathbf{z}^t), 0\} + p_t \max\{-x_{t+1}(\mathbf{z}^t), 0\}).$$

The optimization problem to be solved is given by the following, two-variant formulation where the minimized quantity is the worst-case value or the worst-case expectation of the objective function:

$$\begin{aligned} & \min_{\mathbf{q}(\mathbf{z}), u} u \\ & \text{s.t. } \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \text{ or } \sup_{\mathbf{z} \in \mathcal{Z}} \left\{ -s \max\{x_{T+1}(\mathbf{z}^T), 0\} + \right. \\ & \quad \left. + \sum_{t=1}^T (c_t q_t(\mathbf{z}^{t-1}) + h_t \max\{x_{t+1}(\mathbf{z}^t), 0\} + p_t \max\{-x_{t+1}(\mathbf{z}^t), 0\}) \right\} \leq u \\ & \quad x_{t+1}(\mathbf{z}^t) = x_t(\mathbf{z}^{t-1}) + q_t - z_t, \quad t = 1, \dots, T \\ & \quad L_t \leq q_t(\mathbf{z}^{t-1}) \leq U_t, \quad t = 1, \dots, T \\ & \quad \hat{L}_t \leq \sum_{\tau=1}^t q_\tau(\mathbf{z}^{\tau-1}) \leq \hat{U}_t, \quad t = 1, \dots, T, \end{aligned} \tag{31}$$



**Table 1** Ranges for parameter sampling in the inventory experiment.

Parameter	Range	Parameter	Range
$a_t$	$[0, 20]$	$x_1$	$[20, 50]$
$b_t$	$[a_t, a_t + 100]$	$L_t$	0
$c_t, p_t$	$[0, 10]$	$U_t$	$[50, 70]$
$h_t$	$[0, 5]$	$\hat{L}_t$	0
$s$	0	$\hat{U}_t$	$0.8 \sum_{t=1}^T U_t$

where  $\mathcal{Z}$  is the uncertainty set for  $\mathbf{z}$  and  $\mathcal{P}$  is the ambiguity set of probability distributions with support being a subset of  $\mathcal{Z}$ .

We assume that the uncertainty set  $\mathcal{Z}$  is  $\mathcal{Z} = \mathcal{Z}_1 \times \dots \times \mathcal{Z}_T$ , where  $\mathcal{Z}_t = [a_t, b_t]$ ,  $t = 1, \dots, T$ , which corresponds to  $\mathbf{z}$  being a random variable with independent components. We assume that  $\mathbb{E}_{\mathbb{P}} \mu_t = \frac{a_t + b_t}{2}$ , and that  $\mathbb{E}_{\mathbb{P}} |z_t - \mu_t| = \theta(b_t - a_t)$ , yielding the following ambiguity set:

$$\mathcal{P}_{(\mu, d)} = \{\mathbb{P} : \text{supp}(\mathbb{P}) \subset [\mathbf{a}, \mathbf{b}], \quad \mathbb{E}_{\mathbb{P}} \mathbf{z} = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}} |\mathbf{z} - \boldsymbol{\mu}| = \mathbf{d}, \quad z_i \perp z_j \quad \forall i \neq j\},$$

where  $\mathbf{a} = (a_1, \dots, a_T)^T$ ,  $\mathbf{b} = (b_1, \dots, b_T)^T$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_T)^T$ ,  $\mathbf{d} = (d_1, \dots, d_T)^T$ . The ordering decisions are assumed to be linear functions (LDR) of the past demand:  $q_{t+1}(\mathbf{z}^t) = q_{t+1,0} + \sum_{j=1}^t q_{t+1,j} z_j$  and require that  $q_{t+1}(\mathbf{z}^t) \geq 0$  for all  $\mathbf{z} \in \mathcal{Z}$ , for  $t = 2, \dots, T+1$ . We solve the following two variants of problem (31):

- RO solution - the objective function in (31) is preceded by  $\sup_{\mathbf{z} \in \mathcal{Z}}$ ,
- $(\mu, d)$  solution - the objective function in (31) is preceded by  $\sup_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}}$ .

We conduct an experiment with  $T = 6$  and 50 problem instances. We set  $\theta = 0.25$ , corresponding to the mean absolute deviation of the uniform distribution. The ranges for the uniform sampling of parameters are given in Table 1.

**Upper and lower bounds for the expectation of the objective function.** We consider now the first research question of this section. For each inventory problem instance and the optimal solution  $\bar{\mathbf{q}}(\mathbf{z})$ , we compute the following quantities:

- the worst-case expected cost under  $(\mu, d)$  information:  $\sup_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(\bar{\mathbf{q}}(\mathbf{z}), \mathbf{z})$
- the best-case expected cost  $\inf_{\mathbb{P} \in \mathcal{P}_{(\mu, d, \beta)}} \mathbb{E}_{\mathbb{P}} f(\bar{\mathbf{q}}(\mathbf{z}), \mathbf{z})$  with three possibilities for the skewness of the demand distribution, i.e., with  $\beta_t = \beta \in \{0.25, 0.5, 0.75\}$ , corresponding to left-skewness, symmetry, and right-skewness of the demand distribution in all periods, respectively.

The two values provide us with information about the interval within which the expected objective function value lies under three different assumptions on the parameter  $\beta_t$ . Additionally, for each solution we compute the worst-case cost  $\sup_{\mathbf{z} \in \mathcal{Z}} f(\bar{\mathbf{q}}(\mathbf{z}), \mathbf{z})$  to verify how the minimization of the worst-case expectation affects the worst-case performance of the solution.

Table 2 presents the results. As can be expected, the RO solution yields the best worst-case value of 1950 which is far better than the  $(\mu, d)$  solution, whose worst-case value is 2384. Rows 2 to 4 show

**Table 2** Results of the inventory management - worst-case costs and ranges for the expectation of the objective over  $\mathcal{P}_{(\mu,d,\beta)}$ . All numbers are averages.

Objective type	$\beta$	Minimum cost	
		RO	$(\mu, d)$
Worst-case value	-	1950	2384
Expectation range	0.25	[1255,1280]	[1004,1049]
Expectation range	0.5	[1223,1280]	[970,1049]
Expectation range	0.75	[1230,1280]	[994,1049]

**Table 3** Simulation results for the first inventory problem. Numbers in brackets denote the % change compared to the RO solution.

Objective type	Demand sample type	Cost	
		RO	$(\mu, d)$
Objective mean	Uniform sample	<b>1230</b>	<b>994 (-19.6%)</b>
Objective standard deviation	Uniform sample	157	259 (+65%)
Objective mean	$(\mu, d)$ sample	<b>1246</b>	<b>1003 (-19.5%)</b>
Objective standard deviation	$(\mu, d)$ sample	160	265 (+65.6%)

that the  $(\mu, d)$  solution yields not only better upper bounds on the expected value of the solution, but leads also to an improvement of the best-case expectation for all  $\beta$ . For example, for  $\beta = 0.5$  the interval for the expected cost related to the RO solution is given by [1255,1280], whereas for the  $(\mu, d)$  solution it is [970,1049]. That means that the worst-case expectation obtained by the  $(\mu, d)$  solution is better than the worst-case expectation obtained by the RO solution.

**Simulation results.** We now answer the second research question by conducting a simulation study. Since the solutions are obtained with different objective functions, comparing their average-case performance  $\hat{\mathbf{z}}$  in a ‘fair’ way is difficult. We compare their performance on two samples of demand vectors  $\hat{\mathbf{z}}$ :

- uniform sample - demand scenarios  $\hat{\mathbf{z}}$  are sampled from a uniform distribution on  $\mathcal{Z}$ . This sampling method is favorable to RO solutions since no knowledge is assumed about the demand distribution.
- $(\mu, d)$  sample - demand scenarios  $\hat{\mathbf{z}}$  are sampled from a distribution  $\hat{\mathbb{P}} \in \mathcal{P}_{(\mu,d)}$ . That is, first, a discretized distribution  $\hat{\mathbb{P}} \in \mathcal{P}_{(\mu,d)}$  is sampled using the hit-and-run method (Smith 1984). Then, a vector  $\hat{\mathbf{z}}$  is sampled from the distribution  $\hat{\mathbb{P}}$ . This sampling method is favorable to the  $(\mu, d)$  solutions as it provides a sample consistent with the optimization problem variant.

For each instance, we sample  $10^4$  demand scenarios, with both of the sampling methods. Table 3 presents the results. The averages of the objective function values over the two sample types over all instance are put in bold. For example (row 1), the  $(\mu, d)$  solutions perform better on average on the uniform sample, with values 994 and 1230, respectively. A similar observation holds for the

$(\mu, d)$  sample (row 3). On the other hand, the RO solution has the advantage of smaller standard deviations of the objective function values, for example (row 4), 160 compared to 265 for the  $(\mu, d)$  solution. Taking into account the differences in the standard deviations of the two solutions, one can see that the  $(\mu, d)$  solutions are superior to the RO solutions within deviation of one standard deviation. Thus, we conclude that the  $(\mu, d)$  solutions are superior to the linearly adjustable RO solution in terms of their average case performance in both samples used.

### 3.7. Application 2: Inventory management - enhancement of RO solutions

With the good average-case performance of the  $(\mu, d)$  solutions in the previous experiment, we investigate now the following question: can the  $(\mu, d)$  method be used to enhance the average-case performance of RO solutions? That is, is it possible, in case there are multiple optimal solutions to the RO problem, to find the worst-case optimal solution that have a better average cost than the initial worst-case optimal solution? To verify this, for each of the problem instances of the previous subsection we apply the two-step procedure of Section 3.5.

We consider two enhancement types, corresponding to two different objective functions:

- $(\mu, d)$  enhancement:  $\min \sup_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{q}(\mathbf{z}), \mathbf{z})$ ,
- sample enhancement:  $\min \frac{1}{S} \sum_{j=1}^S f(\mathbf{q}(\hat{\mathbf{z}}_j), \hat{\mathbf{z}}_j)$ , where  $\hat{\mathbf{z}}_j$  are  $S = 200$  demand scenarios sampled uniformly from  $\mathcal{Z}$ .

Table 4 presents the results. On the uniform sample (row 1) the  $(\mu, d)$ -enhanced solution yields an average cost of 1168, compared to 1242 for the non-enhanced solution, that is 5.95% less. For the  $(\mu, d)$  sample (row 3) the corresponding number is 5.93%.

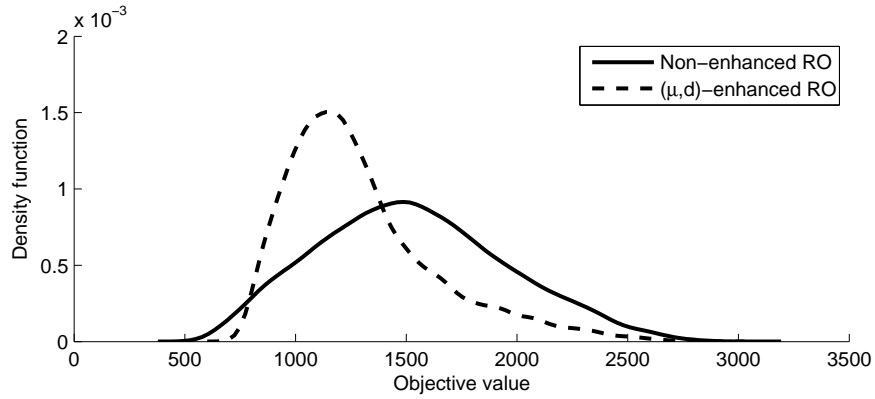
Enhancing the average-case performance of a robust solution does not seem to have a negative effect on the standard deviation of the cost over demand samples. In our example we observe the two types of enhancement,  $(\mu, d)$  enhancement and sample enhancement, to yield nearly the same results. However, the sample method in general problems can provide different solutions for different samples or may need large sample sizes to provide a stable tool for enhancement. We can thus conclude that  $(\mu, d)$  results provide a good and stable tool for enhancing RO solutions when partial distributional information is available. To illustrate the usefulness of the enhancement step, we present also a plot (Figure 1) of kernel density estimates of the cost distribution obtained by the non-enhanced and  $(\mu, d)$ -enhanced solutions in a sample problem on the uniform demand sample.<sup>4</sup> As can be seen, the  $(\mu, d)$  enhancement shifts the density function of the simulated cost significantly to the left, compared to the non-enhanced RO solution. Thus, the average cost becomes smaller and the probability of the cost being less than a given amount becomes larger than for the initial RO solution.

<sup>4</sup> Plot has been obtained using the default normal kernel of the `ksdensity()` function in MATLAB and the corresponding optimal bandwidth.

**Table 4** Results of the inventory management - enhancement of RO solutions example. All numbers are averages. Numbers in brackets denote the % change compared to the initial solution with no enhancement (first column).

Objective type	Enhancement type	Cost		
		-	$(\mu, d)$	Sample
Objective mean	Uniform sample	<b>1242</b>	<b>1168 (-5.95%)</b>	1168 (-5.95%)
Objective standard deviation	Uniform sample	157	158 (+0.63%)	156 (-0.63%)
Objective mean	$(\mu, d)$ sample	<b>1246</b>	<b>1172 (-5.93%)</b>	1172 (-5.93%)
Objective standard deviation	$(\mu, d)$ sample	160	161 (+0.62%)	160 (0.00%)

**Figure 1** Single problem instance. Kernel density estimates of the cost in simulation on the uniform demand sample for the non-enhanced RO solution and the  $(\mu, d)$ -enhanced RO solution.



### 3.8. Application 3: Synthesis of antenna array

We now illustrate the use of the  $(\mu, d)$  results in the context of incorporation of the implementation error in problems with nonlinear constraints. We consider the antenna design problem from Section 7.1.2 of Ben-Tal et al. (2009).

In this problem,  $n$  harmonic oscillators are placed at the points  $k\mathbf{i}$ ,  $k = 1, \dots, n$ , with  $\mathbf{i}$  being the unit vector in the direction of the  $x$ -axis in  $\mathbb{R}^3$ . The objective is to concentrate the energy sent by the antennas within a certain region of the 3-D space, defined using the angle that 3-D directions make with the  $x$  axis.

The electric field generated by an antenna are characterized by the *diagram* of an antenna. The diagram of the  $k$ -th antenna sent in direction  $e$  is given by:

$$D_k(\phi) = \exp(2\pi\imath \cos(\phi)/\lambda),$$

where  $\phi$  is the angle between direction  $e$  and the direction  $\mathbf{i}$  of the  $X$ -axis,  $\lambda$  is the wavelength, and  $\imath$  is the imaginary unit. With complex weights vector  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{C}^n$ , the diagram of the antenna array is the sum of diagrams of the antennas:

$$D(\phi) = \sum_{k=1}^n x_k D_k(\phi).$$

Energy sent by an antenna in the direction given by an angle  $\phi$  from the  $x$ -axis is proportional to the 2-norm of the diagram. The objective is to send as much energy as possible into the region  $\phi \in [0, \Delta]$  by minimizing the weighted  $L_2$  norm of the diagram  $D(\cdot)$  in the sidelobe angle ( $\mathcal{SA}$ )  $\Delta \leq \phi \leq \pi$ :

$$\|D(\cdot)\|_{\mathcal{SA}} = \left( \frac{1}{1 + \cos(\Delta)} \int_{\Delta}^{\pi} |D(\phi)| \sin(\phi) d\phi \right)^{1/2}.$$

The quantity  $\|D(\cdot)\|_{\mathcal{SA}}$  can also be formulated as  $\|\mathbf{Ax}\|$  where  $\mathbf{A} \in \mathbb{C}^{n \times n}$  is such that

$$\mathbf{A} = \mathbf{H}^{1/2}, \quad \mathbf{H} \in \mathbb{C}^{n \times n} : H_{pq} = \frac{1}{1 + \cos \Delta} \int_{\Delta}^{\pi} D_p(\phi) \overline{D_q(\phi)} \sin(\phi) d\phi.$$

For the problem to be bounded, a normalization restriction is added:  $\Re(D(0)) \geq 0$ . Weights  $x_k$  represent the electric power sent to each of the antennas and as such, are subject to implementation error. We assume that the weights  $x_k$  are distorted by the relative implementation error  $\eta_k \in \mathbb{C}$  in the following fashion:

$$x_k \mapsto (1 + \eta_k)x_k.$$

We assume that the real and imaginary parts of the implementation error are independent random variables with supports included in the interval  $[-\rho, \rho]$ , with mean 0 and MAD equal to  $\theta\rho$ :

$$\begin{aligned} \mathcal{P} &= \{\mathbb{P} : \text{supp}(\Im(\eta_k)), \text{supp}(\Re(\eta_k)) \subset [-\rho, \rho], \quad \mathbb{E}_{\mathbb{P}} \Re(\eta_k) = \mathbb{E}_{\mathbb{P}} \Im(\eta_k) = 0, \\ &\quad \mathbb{E}_{\mathbb{P}} |\Im(\eta_k)| = \mathbb{E}_{\mathbb{P}} |\Re(\eta_k)| = \theta\rho, \quad \Im(\eta_k) \perp \Re(\eta_k), \quad k = 1, \dots, n\}. \end{aligned}$$

For the problem to be bounded, a normalization restriction is added:  $\Re(D(0)) \geq 0$ . The optimization problem is given by:

$$\begin{aligned} \min_{\tau, \mathbf{x}} \quad & \tau \\ \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \|\mathbf{Ax}(\boldsymbol{\eta})\| \leq \tau \\ & \Re \left( \sum_{k=1}^n x_k (1 + \eta_k) D_k(0) \right) \geq 1, \quad \forall \boldsymbol{\eta} \in \text{supp}(\boldsymbol{\eta}), \end{aligned} \tag{32}$$

where  $\mathbf{x}(\boldsymbol{\eta}) = [x_1(1 + \eta_1), \dots, x_n(1 + \eta_n)]^T$ ,  $\text{supp}(\boldsymbol{\eta}) = \text{supp}(\eta_1) \times \dots \times \text{supp}(\eta_n)$ . The second constraint in the problem can be reformulated as a deterministic constraint:

$$\Re \left( \sum_{k=1}^n x_k D_k(0) \right) \geq 1 + \rho \sum_{k=1}^n |\Re(x_k D_k(0))| + \rho \sum_{k=1}^n |\Im(x_k D_k(0))|.$$

We solve problem (32) with  $n = 5$  antennas, wavelength  $\lambda = 8$  and  $\Delta = \pi/6$  in two ways:

- **nominal:** in this case we assume  $\rho = 0$  (no implementation error)
- **robust:** we assume  $\rho = 0.01$  (that is, implementation error of 1%) and  $\theta = 0.5$ .

**Table 5** Results of the antenna design experiment. The numbers in the columns are the mean values of simulated  $\|D(\cdot)\|_{\mathcal{SA}}$  (to be minimized in the optimization problem). The numbers in brackets are standard deviations.

Solution	Simulated $\ D(\cdot)\ _{\mathcal{SA}}$	
	Nominal	Robust
$\hat{\rho} = 0$	<b>0.204 (0.00)</b>	<b>0.260 (0.00)</b>
$\hat{\rho} = 0.01$	<b>0.424 (0.19)</b>	<b>0.262 (0.00)</b>
$\hat{\rho} = 0.03$	1.107 (1.41)	0.278 (0.01)
$\hat{\rho} = 0.05$	1.223 (1.32)	0.308 (0.03)
$\hat{\rho} = 0.1$	1.277 (1.78)	0.424 (0.13)

To compare the nominal and robust solutions, we sample uniformly  $10^4$  random perturbations  $\hat{\boldsymbol{\eta}}$  from the set  $\mathcal{E}(\rho) = \{\boldsymbol{\eta} : -\rho \leq \Re(\eta_k), \Im(\eta_k) \leq \rho, \quad k = 1, \dots, n\}$ , with  $\rho \in \{0.01, 0.03, 0.05, 0.1\}$  and compute the value  $\|D(\cdot)\|_{\mathcal{SA}}$  for  $\mathbf{x}(\hat{\boldsymbol{\eta}})$ . Since the normalization condition may not hold with perturbation, we normalize the diagrams  $D(\cdot)$  in such a way that  $|D(0)| = 1$ . Table 5 presents the results.

As visible, the nominal solution performs well only in case of no implementation error, yielding an average value of 0.204, compared to 0.260 for the robust solution. However, already with the relative implementation error equal to 1%, the robust solution performs significantly better, yielding an average value 0.262 (st. dev. 0.0016), compared to 0.424 (0.19) for the nominal solution. This relationship grows even bigger for larger error values, compare 1.277 (1.78) to 0.424 (0.13) in case of 10% relative implementation error. This illustrates that the  $(\mu, d)$  results provide a good way of tackling the implementation error in nonlinear constraints in a distributionally robust way.

## 4. Safe approximations of chance constraints

### 4.1. Introduction

In this section we show how the results of Ben-Tal and Hochman (1972) can be used to construct safe tractable approximations of scalar chance constraints:

$$\mathbb{P}(\mathbf{a}^T(\mathbf{z})\mathbf{x} > \mathbf{b}(\mathbf{z})) \leq \epsilon, \quad \forall \mathbb{P} \in \mathcal{P}_{(\mu, d)}, \quad \text{where } [\mathbf{a}(\mathbf{z}); \mathbf{b}(\mathbf{z})] = [\mathbf{a}^0; \mathbf{b}^0] + \sum_{i=1}^{n_{\mathbf{z}}} z_i [\mathbf{a}_i^0; \mathbf{b}_i^0]. \quad (33)$$

Without loss of generality we assume that the components  $z_i$  have a support contained in  $[-1, 1]$  and mean 0:

$$\mathcal{P}_{(\mu, d)} = \{\mathbb{P} : \text{supp}(z_i) \subseteq [-1, 1], \quad \mathbb{E}_{\mathbb{P}} z_i = 0, \quad \mathbb{E}_{\mathbb{P}} |z_i| = d_i, \quad i = 1, \dots, n_{\mathbf{z}}, \quad z_i \perp z_j, \quad \forall i \neq j\}.$$

To construct the safe tractable approximations, we use the mathematical framework of Ben-Tal et al. (2009). In this framework, the key step consists of bounding from the above the *moment-generating function* of  $z_i, i = 1, \dots, n_{\mathbf{z}}$ :

$$\mathbb{E}_{\mathbb{P}} \exp(wz_i) = \int \exp(wz_i) d\mathbb{P}_i(z_i)$$

and then, using the resulting bound in combination with the Markov inequality to obtain upper bounds on the probability  $\mathbb{P}(\mathbf{a}^T(\mathbf{z})\mathbf{x} > \mathbf{b}(\mathbf{z}))$ .

A strong motivation for using the ambiguity set  $\mathcal{P}_{(\mu,d)}$  is due to the fact that a tight explicit bound on  $\mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T \mathbf{z})$  is obtained easily in this setting by the (BH) results described in Section 2. Indeed, due to the independence of  $z_1, \dots, z_{n_{\mathbf{z}}}$  we have:

$$\begin{aligned} \sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \exp(\mathbf{z}^T \mathbf{w}) &= \prod_{i=1}^{n_{\mathbf{z}}} \sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \exp(z_i w_i) \\ &= \prod_{i=1}^{n_{\mathbf{z}}} \left( \frac{d_i}{2} \exp(-w_i) + \frac{d_i}{2} \exp(w_i) + (1 - d_i) \exp(0) \right) \\ &= \prod_{i=1}^{n_{\mathbf{z}}} (d_i \cosh(w_i) + 1 - d_i). \end{aligned} \quad (34)$$

Notice that the worst-case expectation is evaluated separately per each component of  $\mathbf{z}$ , avoiding the computational burden of summation of  $3^{n_{\mathbf{z}}}$  terms as in (14). In Appendix B we show that in the setting of Wiesemann et al. (2014) without independence of  $z_i$ 's, obtaining the tight upper bound on  $\exp(\mathbf{w}^T \mathbf{z})$  requires solving an optimization problem involving uncertain constraint on a convex function. This requires an exponential number of constraints for an exact reformulation.

#### 4.2. Safe approximations - results

We now show how (34) can be used to obtain safe approximations of (33). First, we present two simple safe approximations in order of increasing tightness. Later, we show that the  $(\mu, d)$  information is particularly suitable for obtaining even tighter safe approximations, based on the exponential polynomials.

The first approximation requires the use of Theorem 2.4.4 of Ben-Tal et al. (2009), repeated in Appendix A.1.

**Theorem 1** *If for a given vector  $\mathbf{x}$  there exist  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n_{\mathbf{z}}+1}$  such that  $(\mathbf{x}, \mathbf{u}, \mathbf{v})$  satisfies the constraint system*

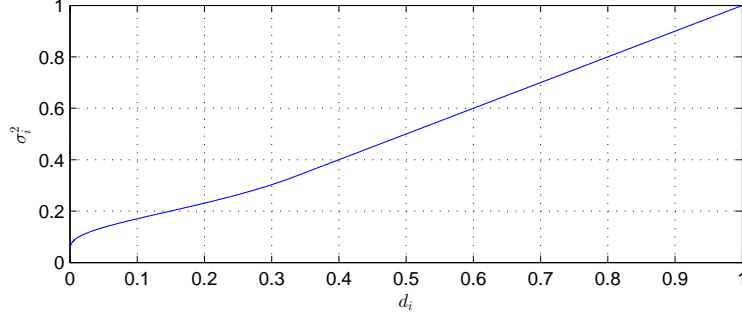
$$\begin{cases} (\mathbf{a}^i)^T \mathbf{x} - b_i = u_i + v_i, 0 \leq i \leq n_{\mathbf{z}} \\ u_0 + \sum_{i=1}^{n_{\mathbf{z}}} |u_i| \leq 0 \\ v_0 + \sqrt{2 \log(1/\epsilon)} \sqrt{\sum_{i=1}^{n_{\mathbf{z}}} \sigma_i^2 v_i^2} \leq 0, \end{cases} \quad (35)$$

where

$$\sigma_i = \sup_{t \in \mathbb{R}} \sqrt{\frac{2 \log(d_i \cosh(t) + 1 - d_i)}{t^2}}, \quad (36)$$

then  $\mathbf{x}$  is feasible to (33), that is, constraint system (35) is a safe approximation of (33). Moreover, (35) is the robust counterpart of the following robust constraint

$$\mathbf{a}^T(\mathbf{z})\mathbf{x} \leq \mathbf{b}(\mathbf{z}), \quad \forall \mathbf{z} \in \mathcal{U}, \text{ where } [\mathbf{a}(\mathbf{z}); \mathbf{b}(\mathbf{z})] = [\mathbf{a}^0; \mathbf{b}^0] + \sum_{i=1}^{n_{\mathbf{z}}} z_i [\mathbf{a}_i^0; \mathbf{b}_i^0], \quad (37)$$

**Figure 2** Plot of  $\sigma_i^2$  as a function of  $d_i$ .

and

$$\mathcal{U} = \left\{ \mathbf{z} \in \mathbb{R}^{n_{\mathbf{z}}} : \exists \mathbf{u} \in \mathbb{R}^{n_{\mathbf{z}}} : \begin{array}{l} z_i - u_i = 0, \quad i = 1, \dots, n_{\mathbf{z}} \\ \sqrt{\sum_{i=1}^{n_{\mathbf{z}}} \frac{u_i^2}{\sigma_i^2}} \leq \sqrt{2 \log(1/\epsilon)} \\ -1 \leq z_i \leq 1, \quad i = 1, \dots, n_{\mathbf{z}} \end{array} \right\}.$$

*Proof.* The proof follows the steps leading to Theorem 2.4.4 from Ben-Tal et al. (2009). First, we need to find scalars  $\mu_i^-, \mu_i^+, \sigma_i$ , where  $i = 1, \dots, n_{\mathbf{z}}$  such that:

$$\int_{-1}^1 \exp(tz_i) d\mathbb{P}_i(z_i) \leq \exp\left(\max\{\mu_i^-, \mu_i^+\} + \frac{\sigma_i^2}{2}\right), \quad \forall t \in \mathbb{R}, \quad \forall \mathbb{P} \in \mathcal{P}_{(\mu, d)}.$$

By (34) we have  $\sup_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \left\{ \int_{-1}^1 \exp(tz_i) d\mathbb{P}_i(z_i) \right\} = d_i \cosh(t) + 1 - d_i$ . Thus, for each  $i$  we need to find  $\mu_i^-, \mu_i^+, \sigma_i$  such that:

$$d_i \cosh(t) + 1 - d_i \leq \exp\left(\max\{\mu_i^- t, \mu_i^+ t\} + \frac{\sigma_i^2 t^2}{2}\right), \quad \forall t \in \mathbb{R}.$$

Setting  $\mu_i^- = \mu_i^+ = 0$ , we then need  $\sigma_i$  such that

$$\begin{aligned} d_i \cosh(t) + 1 - d_i &\leq \exp\left(\frac{\sigma_i^2 t^2}{2}\right), \quad \forall t \in \mathbb{R} \quad \Leftrightarrow \\ \Leftrightarrow \quad \sigma_i^2 &\geq g_i(t) = \frac{2}{t^2} \log(d_i \cosh(t) + 1 - d_i), \quad \forall t \in \mathbb{R}. \end{aligned}$$

Thus, we look for the maximum value of  $g_i(t)$  over the real axis. From the definition of  $g_i(t)$  we know that it is finite, nonnegative, and differentiable everywhere apart from 0. By de l'Hôpital rule we have that  $\lim_{t \rightarrow 0} g_i(t) = d_i$ . It holds that  $\lim_{t \rightarrow \pm\infty} g_i(t) = 0$ . Value of  $\sigma_i$  can be obtained by means of numerical analysis. Figure 2 presents the plot of  $\sigma_i^2$  as a function of  $d_i$ .

From here, by inserting the values  $\mu_i^- = \mu_i^+ = 0, \sigma_i$  into Theorem 2.4.4 of Ben-Tal et al. (2009) (see Appendix A.1), we obtain that robust constraint (37) with  $\mathcal{U}$  defined as above, is a safe tractable approximation of chance constraint (33). By the same theorem, it holds that (35) is precisely the robust counterpart of (37).  $\square$



Constraint system (35) involves only linear and second-order conic constraints, making it highly tractable even for large-dimensional problems.

The second safe approximation is tighter and relies on the somewhat more complicated mathematical machinery of Ben-Tal et al. (2009).

**Theorem 2** *If there exists  $\alpha > 0$  such that  $(\mathbf{x}, \alpha)$  satisfies the constraint*

$$(\mathbf{a}^0)^T \mathbf{x} - b_0 + \alpha \log \left( \sum_{i=1}^{n_{\mathbf{z}}} \left( d_i \cosh \left( \frac{(\mathbf{a}^i)^T \mathbf{x} - b_i}{\alpha} \right) + 1 - d_i \right) \right) + \alpha \log(1/\epsilon) \leq 0, \quad (38)$$

*then  $\mathbf{x}$  satisfies constraint (33). That, is (38) is a safe approximation of (33).*

*Proof.* See Appendix A.2. □

Similar to Theorem 1, one can construct an explicit convex uncertainty set  $\mathcal{U}$  for which (38) is the robust counterpart of (37) corresponding to  $\mathcal{U}$ . Constraint (38) is convex in  $(\mathbf{x}, \alpha)$ , being a sum of a linear function and  $n_{\mathbf{z}}$  perspective functions of the convex log-sum-exp function, see Boyd and Vandenberghe (2004). For that reason, it can be handled with convex optimization algorithms such as Interior Point Methods.

### 4.3. Towards better safe approximations - exponential polynomials

Ben-Tal et al. (2009) discuss the fact that the bounds obtained using a single exponential function can still be improved using the notion of exponential polynomials:

$$\gamma(s) = \sum_{\nu=0}^L c_{\nu} \exp\{\omega_{\nu} s\}, \quad (39)$$

where  $c_{\nu}, \omega_{\nu}, \nu = 0, \dots, L$  are complex numbers and

$$\gamma(\cdot) \text{ is convex and nondecreasing, } \gamma(s) \geq 0, \quad \gamma(0) \geq 0, \quad \gamma(s) \rightarrow 0, \quad s \rightarrow -\infty. \quad (40)$$

In fact, the bound found in Theorem 2 is obtained for a special case where  $L = 0$ ,  $c_0 = \omega_0 = 1$ . The difficulty of using general polynomials (39) lies in the (in)availability of tight, computationally tractable upper bounding function  $\Psi(\mathbf{w})$  on (39):

$$\mathbb{E}_{\mathbb{P}} \gamma \left( w_0 + \sum_{i=1}^{n_{\mathbf{z}}} w_i z_i \right) \leq \Psi(\mathbf{w}), \quad \forall \mathbb{P} \in \mathcal{P}.$$

In the following, we show that under  $(\mu, d)$  information, the result of (BH) can be easily applied also in this case. Indeed, the corresponding supremum over  $\mathcal{P}_{(\mu, d)}$  is given by:

$$\begin{aligned} \Psi(\mathbf{w}) &= \sup_{\mathbb{P} \in \mathcal{P}_{(\mu, d)}} \mathbb{E}_{\mathbb{P}} \gamma \left( w_0 + \sum_{i=1}^{n_{\mathbf{z}}} z_i w_i \right) \\ &= \sum_{\nu=0}^L c_{\nu} \exp\{\omega_{\nu} w_0\} \prod_{i=1}^{n_{\mathbf{z}}} (d_i \sinh(\omega_{\nu} w_i) + 1 - d_i). \end{aligned} \quad (41)$$

Now, we can use Proposition 4.3.1 from Ben-Tal et al. (2009) to obtain the following result.

**Theorem 3** Consider an exponential polynomial  $\gamma(s)$  satisfying (40), the corresponding function  $\Psi(\mathbf{w})$  and the set  $\Gamma_\epsilon$  such that:

$$\Gamma_\epsilon = \{\mathbf{x} : \exists \alpha > 0 : \Psi(\alpha \mathbf{w}) \leq \epsilon\}, \quad w_i = (\mathbf{a}^i)^T \mathbf{x} - b_i, \quad i = 1, \dots, n_{\mathbf{z}}. \quad (42)$$

Then, any  $\mathbf{x} \in \text{cl}\Gamma_\epsilon$  is also feasible for the chance constraint (33).

*Proof.* See Appendix A.3. □

It is also important to note that constraint (42) is convex representable in  $(\mathbf{x}, \alpha)$ . Theorem 3 extends the results of Ben-Tal et al. (2009), which provides a safe approximation using only known supports and means of the components  $z_i$ .

#### 4.4. Safe tractable approximations - simple experiment

We illustrate now the differences between (i) the power of the three approximations of the previous sections, and (ii) knowing and not knowing the MAD. We consider here the following problem from Section 4.3.6.2 of Ben-Tal et al. (2009):

$$\begin{aligned} \max_{x_0} \quad & x_0 \\ \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}(\mu, d)} \mathbb{P} \left( x_0 + \sum_{i=1}^{n_{\mathbf{z}}} x_i z_i > 0 \right) \leq \epsilon \\ & x_i = 1, \quad i = 1, \dots, n_{\mathbf{z}}. \end{aligned} \quad (43)$$

We solve this problem using all three safe tractable approximations of the chance constraint, for two different cases:

- no information about  $d$  - which corresponds to setting  $d_i = 1, i = 1, \dots, n_{\mathbf{z}}$  (the largest possible value for  $d_i$ , see Remark 1, page 5, about the Edmundson-Madansky bound when  $d$  is maximum possible),
- knowing that  $d_i = d = 0.5, i = 1, \dots, n_{\mathbf{z}}$ .

We consider three probability levels:  $\epsilon \in \{10^{-1}, 10^{-2}, 10^{-3}\}$ . Whereas safe approximations corresponding to Theorems 1 and 2 are well-defined by the theorems, we need to choose the exponential polynomial used in the approximation of Theorem 3. As Ben-Tal et al. (2009), we use the polynomial

$$\gamma_{d,T}(s) = \exp(s) \chi_{c^*}(s),$$

where

$$\chi_{c^*}(s) = \sum_{\nu=0}^d (c_\nu^* \exp(i\pi\nu s/T) + \overline{c_\nu^*} \exp(-i\pi\nu s/T))$$

is an optimal solution of the best uniform approximation problem:

$$\mathbf{c}^* \in \arg \min \left\{ \max_{-T \leq s \leq T} |\exp(s) \chi_{\mathbf{c}}(s) - \max\{1 + s, 0\}| : 0 \leq \chi_{\mathbf{c}}(s) \leq \chi_{\mathbf{c}}(0) = 1, \quad \forall s \in \mathbb{R} \right\}$$

**Table 6** Maximum values of  $x_0$  in problem (43), depending on the safe tractable approximation used, probability bound, and the assumptions on the knowledge about  $d$ .

$\epsilon$		Maximum $x_0$		
	Safe approximation	Theorem 1	Theorem 2	Theorem 3
$10^{-1}$	Unknown $d$	-24.28	-24.21	-20.43
	$d = 0.5$	-17.16	-17.14	-14.48
$10^{-2}$	Unknown $d$	-34.34	-34.13	-30.55
	$d = 0.5$	-24.27	-24.20	-21.69
$10^{-3}$	Unknown $d$	-42.05	-41.67	-38.34
	$d = 0.5$	-29.73	-29.60	-27.25

and  $\exp(s)\chi_\epsilon(s)$  is convex nondecreasing on  $[-T, T]$ , with parameter values  $d = 11, T = 8$ .

Table 6 presents the results. First, for all safe approximations and all security levels, one can observe a substantial value of having the information about parameters  $d_i$ . For example, for  $\epsilon = 0.01$  and safe approximation according to Theorem 3, the optimal solution obtained without knowledge of  $d$  is  $-30.55$ , whereas the corresponding number for known  $d = 0.5$  is  $-21.69$ . Similar pattern can be observed for other values of  $\epsilon$  and other approximations.

Secondly, one can see the increasing power of the safe tractable approximations that use exactly the same information. For example, for  $\epsilon = 10^{-3}$  and  $d = 0.5$  the subsequent optimal values are  $-29.73$ ,  $-29.60$  and  $-27.25$ . For all values of  $\epsilon$  and  $d$  there is a bigger difference between the second and third tractable approximation than between the first and second.

This example illustrates the extra power given by the knowledge of  $d$  and gives a strong reason to estimate this quantity in order to obtain better chance constraint approximations. Also, the difference between the quality of safe tractable approximations of Theorems 1, 2, and 3 illustrates that the power of exponential polynomial-based approximations make them an attractive tool if the parameters  $a$ ,  $b$ ,  $\mu$ , and  $d$  can be estimated.

#### 4.5. Antenna array - chance constraints

Here we consider an application of the results on safe tractable approximations to scalar chance constraints to an antenna design problem. The setting of the problem is as follows. There are  $n = 40$  ring-shaped antennas belonging to the  $XY$  plane in  $\mathbb{R}^3$ . The radius of the  $k$ -th antenna is defined as  $k/n$  and the diagram  $D(\phi)$  of the antenna array is defined as a sum of diagrams  $D_k(\phi)$  of the antennas:

$$D_k(\phi) = \frac{1}{2} \int_0^{2\pi} \cos\left(\frac{2\pi k}{40} \cos(\phi) \cos(\gamma)\right) d\gamma.$$

The objective of the problem is to minimize the maximum of the diagram modulus in the angle of interest  $0 \leq \phi \leq 70^\circ$ :

$$\max_{0 \leq \phi \leq 70^\circ} \left| \sum_{k=1}^n x_k D_k(\phi) \right|,$$

subject to the restrictions that:

- the diagram in the interval  $77^\circ \leq \phi \leq 90^\circ$  is nearly uniform:

$$77^\circ \leq \phi \leq 90^\circ \quad \Rightarrow \quad 0.9 \leq \sum x_k D_k(\phi) \leq 1,$$

- the diagram in other angles is not too large:

$$\left| \sum_{k=1}^n x_k D_k(\phi) \right| \leq 1, \quad 70^\circ \leq \phi \leq 77^\circ.$$

We assume that the implementation error affects the weight of the  $k$ -th antenna in the following fashion:

$$x_k \mapsto \tilde{x}_k = (1 + z_k \rho) x_k, \quad k = 1, \dots, n,$$

where  $z_k$ ,  $k = 1, \dots, n$ , are independent random variables with supports contained in  $[-1, 1]$ , with mean 0 and MAD equal to  $d$ :

$$\mathcal{P} = \{ \mathbb{P} : \text{supp}(z_i) \subset [-1, 1], \quad \mathbb{E}_{\mathbb{P}}(z_i) = 0, \quad \mathbb{E}_{\mathbb{P}}|z_i| = d, \quad z_i \perp z_j, \quad \forall i \neq j \}.$$

The problem to be solved is:

$$\begin{aligned} & \min_{\tau, \mathbf{x}} \tau \\ & \text{s.t. } \begin{aligned} \mathbb{P}(\sum D_k(\phi_i) \tilde{x}_k \leq \tau) &\geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & 0 \leq \phi_i \leq 70^\circ \\ \mathbb{P}(\sum D_k(\phi_i) \tilde{x}_k \geq -\tau) &\geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & 0 \leq \phi_i < 70^\circ \\ \mathbb{P}(\sum D_k(\phi_i) \tilde{x}_k \leq 1) &\geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & 70^\circ \leq \phi_i < 77^\circ \\ \mathbb{P}(\sum D_k(\phi_i) \tilde{x}_k \geq -1) &\geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & 70^\circ \leq \phi_i \leq 77^\circ \\ \mathbb{P}(\sum D_k(\phi_i) \tilde{x}_k \leq 1) &\geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & 77^\circ \leq \phi_i \leq 90^\circ \\ \mathbb{P}(\sum D_k(\phi_i) \tilde{x}_k \geq 0.9) &\geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & 77^\circ \leq \phi_i \leq 90^\circ, \end{aligned} \end{aligned} \quad (44)$$

where  $\phi_1, \dots, \phi_N$  is a ‘fine grid’ of equidistance placed points on  $[0^\circ, 90^\circ]$ .

In the numerical experiment we set  $N = 400, d = 0.5$ . The chance constraints are reformulated using the ball-box uncertainty set of Theorem 1. We solve the problem in the following settings:

- nominal solution, with  $\rho = 0$  (no implementation error),
- robust solutions with  $(\epsilon, \rho) \in \{0.001, 0.01, 0.05\}^2$ .

In total, we obtain 10 solutions. For each of them we report the optimal (worst-case) objective value. Next to that, we conduct a simulation study for each solution, assuming  $\hat{\rho} \in \{0.001, 0.01, 0.05\}$ . In this study, for each solution we sample  $10^4$  scenarios of the implementation error  $\hat{\mathbf{z}} \in [-1, 1]^n$  and we report on:

**Table 7** Minimum worst-case  $\tau^*$  and mean simulated values of  $\hat{\tau}$  for each of the solutions.  $\epsilon$  and  $\rho$  denote the parameter values used in problem (44) to obtain a given solution, and  $\hat{\rho}$  denotes the error magnitude of the given sample of  $10^4$  implementation error vectors  $\mathbf{z}$ .

$\epsilon$	-	0.001			0.01			0.05			
$\rho$	0	0.001	0.01	0.05	0.001	0.01	0.05	0.001	0.01	0.05	
Worst-case $\tau \times 10^{-2}$	-	2.68	5.86	8.12	34.05	5.77	7.92	16.04	5.68	7.71	11.80
Average $\hat{\tau} \times 10^{-2}$	$\hat{\rho} = 0.001$	4270	5.64	7.66	31.91	5.59	7.53	14.70	5.54	7.38	10.84
	$\hat{\rho} = 0.01$	42706	7.03	7.78	31.91	6.96	7.65	14.70	6.91	7.51	10.85
	$\hat{\rho} = 0.05$	213534	14.00	8.61	31.91	13.73	8.46	14.94	13.64	8.36	11.12

**Table 8** Empirical probabilities of violating at least one constraint. ‘Violation probability (%)’ denotes the percentage of simulated implementation error vectors for which at least one of the constraints of the problem (44) has been violated.

$\epsilon$	-	0.001			0.01			0.05			
$\rho$	0	0.001	0.01	0.05	0.001	0.01	0.05	0.001	0.01	0.05	
Violation probability (%)	$\hat{\rho} = 0.001$	100	0.00	0.00	0.00	0.02	0.00	0.00	0.18	0.00	0.00
	$\hat{\rho} = 0.01$	100	84.39	0.00	0.00	84.87	0.00	0.00	85.67	0.20	0.00
	$\hat{\rho} = 0.05$	100	99.57	63.89	0.00	99.66	62.97	0.03	99.61	67.35	0.43

- the percentage of samples for which at least one of the constraints of the problem (44) is violated,

- the value  $\hat{\tau} = \max_{0^\circ \leq \phi_i \leq 70^\circ} |\sum \hat{x}_k D_k(\phi_i)|$ .

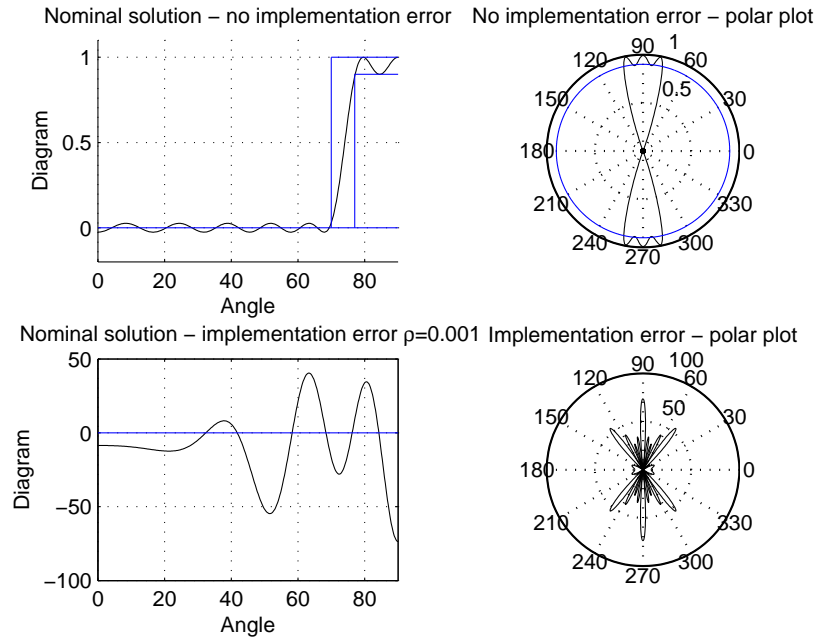
Results are given in Tables 7 and 8. It can be noticed that the nominal solution becomes senseless already with the implementation error  $\rho = 0.001$ . At the same time, the robust solutions do still yield good performance, both in terms of the  $\hat{\tau}$  values and the percentage of samples for which at least one constraint is violated.

The difference between the nominal and robust solutions can be seen in Figures 3 and 4, where the diagrams are plotted for the situations (i) with no implementation error, (ii) with a single sample of implementation error  $\hat{\rho} = 0.001$ . In both cases, solutions yield good ‘desired’ diagrams in the no-error case. However, in the situation with implementation error (lower panels), the robust solution still ‘fits’ into the desired bounds, which is completely not the case for the nominal solution.

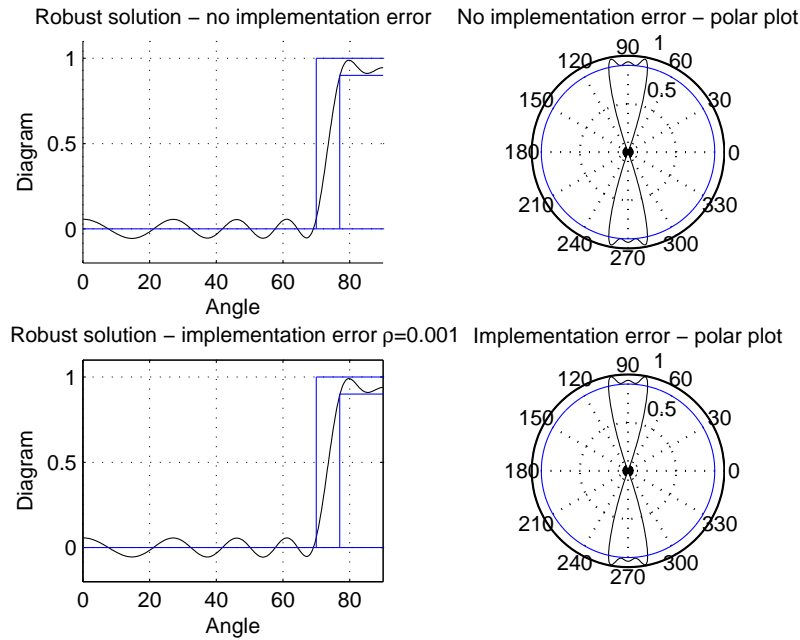
## 5. Summary

In this paper, we have considered two types of ambiguous stochastic constraints - expected feasibility constraints and chance constraints. In contrast to previous research, which employs the variance as a dispersion measure, we use the mean absolute deviation. This allows us to use the 1972 results of (BH) on tight upper and lower bounds on the expectation of a convex function of a

**Figure 3** Nominal solution - diagram plots. Upper panel - situation without implementation error. Lower panel - implementation error, single trajectory.



**Figure 4** Robust solution - diagram plots. Upper panel - situation without implementation error. Lower panel - implementation error, single trajectory.



random variable, and thus, to provide tractable exact robust counterparts for expected feasibility constraint and to obtain safe tractable approximations of ambiguous chance constraint.

To make our approach self-contained, we outline a statistical procedure of estimating the needed and relevant parameters. Numerical examples show the proposed methodology to perform well and, in particular, to offer substantial improvements in the worst-case expected performance and probabilistic guarantees on constraints' feasibility.

## References

- Ben-Tal, A., & Hochman, E. (1972). More bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability*, Vol. 9, pp. 803-812.
- Ben-Tal, A., & Hochman, E. (1976). Stochastic programs with incomplete information. *Operations Research*, Vol. 24(2), pp. 336-347.
- Ben-Tal, A., & Hochman, E. (1985). Approximation of expected returns and optimal decisions under uncertainty using mean and mean absolute deviation. *Zeitschrift für Operations Research*, Vol. 29(7), pp. 285-300.
- Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, Vol. 23(4), pp. 769-805.
- Ben-Tal, A., Goryashko, A., Guslitzer, E., & Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, Vol. 99(2), pp. 351-376.
- Ben-Tal, A., Golany, B., Nemirovski, A., & Vial, J.-Ph. (2005). Retailer-supplier flexible commitments contracts: a robust optimization approach. *Manufacturing & Service Operations Management*, Vol. 7(3), pp. 248-271.
- Ben-Tal, A., El Ghaoui, L. & Nemirovski, A. (2009). *Robust optimization*. Princeton University Press.
- Ben-Tal, A., & Hertog, D. den (2011). Immunizing conic quadratic optimization problems against implementation errors. *Mathematical Programming*, Vol. 149(1-2) pp. 265-299.
- Ben-Tal, A., Hertog, D. den, De Waegenaere, A., Melenberg, B. & Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, Vol. 59(2), pp. 341-357.
- Ben-Tal, A., Hertog, D. den, & Vial, J.-Ph. (2015). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*. Vol. 149(1-2), pp. 265-299.
- Bertsimas, D., Goyal, V., & Sun, X. A. (2011). A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Mathematics of Operations Research*, Vol. 36(1), pp. 24-54.
- Birge, J. R., & Wets, R. J. B. (1987). Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research*, Vol. 12(1), pp. 149-162.
- Birge, J. R., & Dulá, J. H. (1991). Bounding separable recourse functions with limited distribution information. *Annals of Operations Research*, Vol. 30(1), pp. 277-298.

- Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.
- Calafiore, G. C., & El Ghaoui, L. (2006). On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, Vol. 130(1), pp. 1-22.
- Chen, X., Sim, M., Sun, P. (2007) A robust optimization perspective on stochastic programming. *Operations Research*, Vol. 55(6), pp. 1058-1071.
- Chen, W., Sim, M., Sun, J., & Teo, C.-P. (2010) From CVaR to uncertainty set: Implications in joint chance constrained optimization. *Operations Research*, Vol. 58(2) pp. 470-485.
- Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, Vol. 58(3), pp. 595-612.
- Dupačová, J. (1977). Minimax approach to stochastic linear programming and the moment problem (in Czech), *EMO*, Vol. 13, pp. 279-307.
- Edmundson, H. P. (1956). Bounds on the expectation of a convex function of a random variable. *Acta Mathematica*, Vol. 30, pp. 175-193.
- El Amir, E. A. H. (2012). On uses of mean absolute deviation: decomposition, skewness and correlation coefficients. *Metron*, Vol. 70(2-3), pp. 145-164.
- Gallego, G. (1992). A minmax distribution free procedure for the (Q, R) inventory model. *Operations Research Letters*, Vol. 11(1), pp. 55-60.
- Gallego, G., Ryan, J. K., & Simchi-Levi, D. (2001). Minimax analysis for finite-horizon inventory models. *IIE Transactions*, Vol. 33(10), pp. 861-874.
- Gastwirth, J. L. (1974). Large sample theory of some measures of income inequality. *Econometrica: Journal of the Econometric Society*, Vol. 42(1), pp. 191-196.
- Goh, J., & Sim, M. (2010). Distributionally robust optimization and its tractable approximations. *Operations Research*, Vol. 58(4), pp. 902-917.
- Grant, M., & Boyd, S. (2015) CVX: MATLAB software for disciplined convex programming, version 2.0 beta. Available at: <http://cvxr.com/cvx>.
- Hanasusanto, G. A., Roitch, V., Kuhn, D., & Wiesemann, W. (2015). A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*. Available at: <http://link.springer.com/article/10.1007/s10107-015-0896-z>.
- Iancu, D. A., & Trichakis, N. (2013). Pareto efficiency in robust optimization. *Management Science*, Vol. 60(1), pp. 130-147.
- Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, Vol. 30(1), pp. 175-193.
- Jolliffe, I. (2002). *Principal component analysis*. John Wiley & Sons, Ltd.



- Madansky, A. (1959). Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics*, Vol. 30, pp. 743-746.
- Natarajan, K., Pachamanova, D., Sim, M. (2008) Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, Vol. 54(3), pp. 573-585.
- Perakis, G., & Roels, G. (2008). Regret in the newsvendor model with partial information. *Operations Research*, Vol. 56(1), pp. 188-203.
- Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, Vol. 55(1), pp. 98-112.
- Postek, K., Hertog, D. den, & Melenberg, B. (2014) Tractable counterparts of distributionally robust constraints on risk measures. *CentER Discussion Paper Series No. 2014-031*. Available at: [http://www.optimization-online.org/DB\\_HTML/2014/05/4365.html](http://www.optimization-online.org/DB_HTML/2014/05/4365.html).
- Rogosinsky, W. W. (1958). Moments of non-negative mass. *Proceedings of the Royal Society of London, Series A*, Vol. 245, pp. 1-27.
- Ruiter, F. de, Ben-Tal, A., Hertog, D. den, & Brekelmans, R. (2014) Adjustable robust optimization with decision rules based on inexact revealed data. *CentER Discussion Paper Series No. 2014-003*. Available at: [http://www.optimization-online.org/DB\\_FILE/2014/01/4198.pdf](http://www.optimization-online.org/DB_FILE/2014/01/4198.pdf).
- Scarf, H. E. (1958). A min-max solution of an inventory problem. In Arrow, K. J., Karlin, S., & Scarf, H. E., editors. *Studies in the Mathematical Theory of Inventory and Production*, pp. 201-209. Stanford University Press.
- Shapiro, A., Kleywegt, A. J. (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software*, Vol. 17, pp. 523-542.
- Smith, R. L. (1984). Efficient Monte-Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, Vol. 32, pp. 1296-1308.
- Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, Vol. 62(6), pp. 1358-1376.
- Zymler, S., Kuhn, D., & Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, Vol. 137(1-2), pp. 167-198.

## Appendix A: Safe approximations of chance constraints

In this Appendix we list the relevant results from Ben-Tal et al. (2009) used to prove Theorems 1 and 2, and 3 adopted to the notation of this paper.

### A.1. Safe approximation in Theorem 1

In the proof of Theorem 1 the following result is used.

**Theorem 4 (Ben-Tal et al. (2009), Theorem 2.4.4)** *Assume that:*

P.1.  $z_i, i, \dots, n_{\mathbf{z}}$  are independent random variables such that  $\text{supp}(z_i) \subseteq [a_i^-, a_i^+]$ ,  $i = 1, \dots, n_{\mathbf{z}}$ ,

P.2. the distributions  $\mathbb{P}_i$  of the components  $z_i$  are such that

$$\int \exp(ts) d\mathbb{P}_i(s) \leq \exp\left(\max\{\mu_i^+ t, \mu_i^- t\} + \frac{1}{2}\sigma_i^2 t\right), \quad \forall t \in \mathbb{R}, \quad (45)$$

with known constants  $\mu_i^- \leq \mu_i^+$ .

Then, the robust constraint

$$\mathbf{a}^T(\mathbf{z})\mathbf{x} \leq \mathbf{b}(\mathbf{z}), \quad \forall \mathbf{z} \in \mathcal{U}, \text{ where } [\mathbf{a}(\mathbf{z}); \mathbf{b}(\mathbf{z})] = [\mathbf{a}^0; \mathbf{b}^0] + \sum_{i=1}^{n_{\mathbf{z}}} z_i [\mathbf{a}_i^0; \mathbf{b}_i^0], \quad (46)$$

and

$$\mathcal{U} = \left\{ \mathbf{z} \in \mathbb{R}^{n_{\mathbf{z}}} : \exists \mathbf{u} \in \mathbb{R}^{n_{\mathbf{z}}} : \begin{array}{l} \mu_i^- \leq z_i - u_i \leq \mu_i^+, \quad i = 1, \dots, n_{\mathbf{z}} \\ \sqrt{\sum_{i=1}^{n_{\mathbf{z}}} \frac{u_i^2}{\sigma_i^2}} \leq \sqrt{2\log(1/\epsilon)} \\ a_i^- \leq z_i \leq a_i^+, \quad i = 1, \dots, n_{\mathbf{z}} \end{array} \right\},$$

is a safe approximation of (33). Moreover,  $\mathbf{x}$  satisfies (46) if and only if there exist  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n_{\mathbf{z}}+1}$  such that  $\mathbf{x}, \mathbf{u}, \mathbf{v}$  satisfy the following set of constraints:

$$\begin{cases} (\mathbf{a}^i)^T \mathbf{x} - b^i = u_i + v_i, & i = 0, \dots, n_{\mathbf{z}} \\ u_0 + \sum_{i=1}^{n_{\mathbf{z}}} \max\{a_i^+ u_i, a_i^- u_i\} \leq 0 \\ v_0 + \sum_{i=1}^{n_{\mathbf{z}}} \max\{\mu_i^+ v_i, \mu_i^- v_i\} + \sqrt{2\log(1/\epsilon)} \sqrt{\sum_{i=1}^{n_{\mathbf{z}}} \sigma_i^2 v_i^2} \leq 0. \end{cases}$$

## A.2. Safe approximation in Theorem 2

The proof of Theorem 2 relies on the following result from Ben-Tal et al. (2009).

**Theorem 5 (Ben-Tal et al. (2009), Proposition 4.2.2)** Assume that the distribution  $\mathbb{P}$  of the random perturbation  $\mathbf{z}$  is such that

$$\log(\mathbb{E} \exp(\mathbf{w}^T \mathbf{z})) \leq \Phi(\mathbf{w}),$$

where  $\mathbf{w} = (w_1, \dots, w_{n_{\mathbf{z}}})$  for some known convex function  $\Phi(\cdot)$  that is finite everywhere and satisfies  $\Phi(0) = 0$ . Then, any  $(w_0, \mathbf{w})$  feasible for

$$\inf_{\beta > 0} \{w_0 + \beta \Phi(\beta^{-1} \mathbf{w}) + \beta \log(1/\epsilon)\} \leq 0$$

is feasible for the chance constraint

$$\mathbb{P}\left(w_0 + \sum_{i=1}^{n_{\mathbf{z}}} w_i z_i > 0\right) \leq \epsilon.$$

*Proof of Theorem 2.* We show that the function  $\Phi(\mathbf{w})$ :

$$\Phi(\mathbf{w}) = \log(\Psi(\mathbf{w})), \quad \Psi(\mathbf{w}) = \sup_{\mathbb{P} \in \mathcal{P}(\mu, \mathbf{d})} \mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T \mathbf{z}) = \prod_{l=1}^{n_{\mathbf{z}}} (d_l \cosh(w_l) + 1 - d_l).$$

satisfies the conditions of Theorem 5. Indeed, from (BH) we know that  $\Psi(\mathbf{w})$  gives a tight upper bound on  $\mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T \mathbf{z})$ . Also, the function  $\Phi(\mathbf{w})$  is convex as it is the log-sum-exp function, see Boyd and Vandenberghe (2004), and it holds that  $\Phi(0) = 0$ . Thus, it is sufficient to substitute

$$w_i := (\mathbf{a}^i)^T \mathbf{x} - b_i, \quad i = 0, \dots, n_{\mathbf{z}},$$

to arrive at constraint (38) from Theorem 2.  $\square$

### A.3. Safe approximation in Theorem 3

Theorem 3 follows from the following result from Ben-Tal et al. (2009):

**Theorem 6 (Ben-Tal et al. (2009), Proposition 4.3.1)** *Consider a generating function  $\gamma(s)$  satisfying (40). Let  $\Psi(\mathbf{w})$  be a finite convex function satisfying*

$$\Psi(\mathbf{w}) \geq \mathbb{E}_{\mathbb{P}} \left( \gamma \left( w_0 + \sum_{i=1}^{n_{\mathbf{z}}} w_i z_i \right) \right), \quad \Psi(\mathbf{w} + t[-1, 0, \dots, 0]) \rightarrow 0, \text{ when } t \rightarrow \infty.$$

*Then, the inequality*

$$\inf_{\beta > 0} (\beta \Psi(\beta^{-1} \mathbf{w}) - \beta \epsilon) \leq 0$$

*is a safe approximation of the chance constraint*

$$\mathbb{P} \left( w_0 + \sum_{i=1}^{n_{\mathbf{z}}} w_i z_i > 0 \right) \leq \epsilon.$$

*Proof of Theorem 3.* The result follows from using  $\Psi(\mathbf{w})$  defined as in (41). This function clearly satisfies the conditions of Theorem 6. Then, the only remaining part is substituting the relevant terms for  $w_i, i = 0, \dots, n_{\mathbf{z}}$ .  $\square$

## Appendix B: Worst-case expectation of $\exp(\mathbf{w}^T \mathbf{z})$ without independent components

We now consider obtaining an upper bound on  $\exp(\mathbf{w}^T \mathbf{z})$  using the results of Wiesemann et al. (2014), where the components of the random variable  $\mathbf{z}$  are not assumed to be independent. For that reason, the distributional uncertainty set is given by:

$$\mathcal{P}' = \{ \mathbb{P} : \text{supp}(z_i) \subseteq [-1, 1], \quad \mathbb{E}_{\mathbb{P}} z_i = 0, \quad \mathbb{E}_{\mathbb{P}} |z_i| = d_i, \quad i = 1, \dots, n_{\mathbf{z}} \}.$$

To obtain the worst-case expectation, one needs to solve the following problem:

$$\begin{aligned} \min_t & \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}} \exp(\mathbf{z}^T \mathbf{w}) \leq t, \quad \forall \mathbb{P} \in \mathcal{P}' \end{aligned} \tag{47}$$

The uncertainty set for the distributions  $\mathbb{P}$  in their framework is:

$$\mathcal{P}' = \left\{ \mathbb{P} : \mathbb{E}_{\mathbb{P}} \left( \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \mathbf{z} + \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \mathbf{u} \right) = \begin{bmatrix} \mathbf{0} \\ \mathbf{d} \end{bmatrix} \right\}, \quad (48)$$

$$\mathbb{P}((\mathbf{z}, \mathbf{u}) \in \mathcal{C}) = 1$$

where  $\mathcal{C} = \{(\mathbf{z}, \mathbf{u}) : -\mathbf{1} \leq \mathbf{z} \leq \mathbf{1}, \quad \mathbf{u} \geq \mathbf{z}, \quad \mathbf{u} \geq -\mathbf{z}, \quad \mathbf{u} \leq \mathbf{1}\}$ . Then, the problem to solve is equivalent to:

$$\begin{aligned} \min_{\kappa, \lambda \geq 0, \beta_1, \beta_2, t} \quad & t \\ \text{s.t.} \quad & \beta_2^T \mathbf{d} + \mathbf{1}^T (\boldsymbol{\kappa} - \boldsymbol{\lambda}) \leq t \\ & \mathbf{z}^T \beta_1 + \mathbf{u}^T \beta_2 + \mathbf{1}^T (\boldsymbol{\kappa} - \boldsymbol{\lambda}) \geq \exp(\mathbf{z}^T \mathbf{w}), \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C} \end{aligned} \quad (49)$$

The last line of (49) involves a constraint on the function  $\exp(\mathbf{z}^T \mathbf{x})$  over  $\mathcal{C}$ . Since  $\exp(\mathbf{z}^T \mathbf{x})$  is strictly convex in  $\mathbf{z}$ , an equivalent reformulation of such a constraint would have to take into account all  $3^{n_z}$  vertices of  $\mathcal{C}$ . The number  $3^{n_z}$  comes from the fact that per component, the uncertainty set is a triangle  $\mathcal{C}_i = \{(z_i, u_i) : -1 \leq z_i \leq 1, \quad u_i \geq z_i, \quad u_i \geq -z_i, \quad u_i \leq 1\}$ .

### Appendix C: Properties of the MAD

In this Appendix we provide some properties of the MAD and the formulas for several well-known probability distributions, based on Ben-Tal and Hochman (1985).

If we denote by  $\sigma^2$  the variance of the random variable  $z$ , whose distribution is known to belong to the set  $\mathcal{P}_{(\mu, d)}$  (see 6, page 5), then it holds that:

$$\frac{d^2}{4\beta(1-\beta)} \leq \sigma^2 \leq \frac{d(b-a)}{2}.$$

In particular, it holds that  $d \leq \sigma$ . For a proof, we refer the reader to Ben-Tal and Hochman (1985).

For several specific distributions, an explicit formula for  $d$  is available:

- Uniform distribution on  $[a, b]$ :

$$d = \frac{1}{4}(b-a)$$

- Normal distribution  $N(\mu, \sigma^2)$ :

$$d = \sqrt{\frac{2}{\pi}} \sigma$$

- Gamma distribution with parameters  $\lambda$  and  $k$  (for which  $\mu = k/\lambda$ ):

$$d = \frac{2k^k}{\Gamma(k) \exp(k)} \frac{1}{\lambda}.$$

Ben-Tal and Hochman (1985) provide an explicit formula for the MAD for general *stable distributions*. A stable distribution is defined via its *location parameter*  $\kappa$ , *scale parameter*  $D > 0$ , *measure of skewness*  $-1 \leq \lambda \leq 1$ , and *characteristic exponent*  $0 < \alpha \leq 2$ . Important distributions belonging to this class are, for example, the normal and Cauchy distributions. The characteristic function of a stable distribution is given by

$$\log \Psi_z(t) = \log \mathbb{E} \exp(itz) = i\kappa t - D|t|^\alpha \left( 1 + i\lambda \text{sign}(t) \tan \left( \frac{1}{2} \pi \alpha \right) \right).$$

Stable distributions are the only possible limiting laws for sums of independent identically distributed random variables. For properties of the stable distributions we refer the reader to Ben-Tal and Hochman (1985), who prove that for  $1 < \alpha \leq 2$  the MAD of a stable random variable is given by:

$$d = D^{1/\alpha} H(\lambda, \alpha),$$

where

$$H(\lambda, \alpha) = \frac{2}{\pi} \frac{\Gamma(1 - 1/\alpha)}{(1 + A^2)^{(\alpha-1)/2\alpha}} (\cos((1 - 1/\alpha) \arctan A) + A \sin((1 - 1/\alpha) \arctan A)),$$

and  $A = \lambda \tan(\frac{1}{2}\alpha\pi)$ . In case of  $\alpha \leq 1$  the mean of the random variable  $z$  does not exist.